

USER PERCEPTION BENCHMARKS FOR COMMERCIAL AND INSTITUTIONAL BUILDINGS IN NEW ZEALAND

Clare Dykes

Victoria University of Wellington

March 2012

A thesis submitted to the School of Architecture, Victoria University of Wellington, in
fulfilment of the requirements for the degree of Master of Building Science

PREFACE

This research thesis was submitted in fulfilment of the requirements for the degree of Master of Building Science at the School of Architecture, Victoria University of Wellington, New Zealand.

Author:

Clare Dykes

School of Architecture

Victoria University of Wellington

Email: claredykes@gmail.com

Contact Number: +64 27 6298599

Research Supervisor:

Dr George Baird

Professor of Building Science, School of Architecture

Victoria University of Wellington

Email: George.Baird@vuw.ac.nz

ABSTRACT

Buildings are ultimately built for people to use, so it is important that they are designed to suit the occupants' needs. The indoor environment can affect the comfort, health, and productivity of occupants. Yet very little is known about how the building users perceive these matters and there is very little information on what would constitute norms or benchmarks for their perceptions. If we wish to improve users' perceptions then it is essential to have a reliable standardised set of benchmarks. However, at present there is no statistically unbiased, country-based benchmark for the perceptions of the occupants of the building environments that is representative of the existing building stock.

Predominantly, current building performance indicators have the main focus on technical aspects of the building such as energy consumption and materials recycling; very little attention is paid to how the occupants themselves rate the building. User perception benchmarks will allow building owners/users to see how New Zealand buildings are rated by the occupants themselves. Although subjective, it will allow the space to be rated by those who it is designed for, the occupants. This in turn will allow for possible improvements in the construction of comfortable, healthy, and productive work environments.

The overall aim of this study is to contribute to the improvement of the performance of commercial and institutional buildings for their occupants through the exploration of user performance benchmarks. The objectives are to provide a general overview of existing studies which have investigated user perceptions, to explore the nature and characteristics of benchmarks, to create a pilot database of results for user surveys of NZ buildings, and finally to use the data collected to explore methods for determining benchmarks for NZ buildings in terms of user perception.

Using questionnaires as a tool for assessing the performance of a building is an effective way of determining how the building is affecting the occupants, and is cheaper and less time consuming than the technical measures that would be needed to reach the same level of detail. The pilot database allowed flexibility in the statistical analysis and exploration of how possible benchmarks could be constructed. Patterns and influencing factors were investigated in these results.

The results from the occupant questionnaire data from this database followed consistent patterns to those of other established methods, with the *Storage*, *Health* and *Air* variables consistently scoring low. The highest scoring variables are for *Furniture*, *Image*, and *Lighting Overall*. *Image* is the variable with the largest variability, reflecting its subjective nature.

Correlations were carried out on the summary variables, with particular focus on *Health*, *Overall Comfort*, and *Perceived Productivity*. These were highly correlated with each other and many of the other variables. In terms of the indoor environment variables, *Air Overall in Summer*, *Air Overall in Winter*, *Temperature Overall in Summer* and *Temperature Overall in Winter* were consistently the highest correlation between *Health*, *Overall Comfort* and *Perceived Productivity*. *Lighting Overall* had the weakest correlation for all three variables.

It was deemed that it was entirely appropriate and feasible to establish benchmarks from the questionnaire data received, and through preliminary exploration incorporate them into building rating tools. There are a variety of forms these benchmarks can take, with positives and negatives for each.

ACKNOWLEDGMENTS

Many thanks to everyone who has helped during this research process. I would especially like to thank my research supervisor, George Baird, who provided me with great guidance and feedback throughout the year. The valuable resources that he was kind enough to share made this research project possible. I would also like to acknowledge Nigel Isaacs for his willingness to stand-in as my secondary supervisor during George's time overseas.

I would like to acknowledge Michael Donn, Nigel Isaacs and BRANZ for organising financial assistance for this research.

I would also like to acknowledge Adrian Leaman and Building Use Studies (BUS), for the use of his questionnaires, benchmarks, and summary reports.

Many thanks to Patrick Arnold, of E-Cubed Building Workshop, for the use of the building user survey data that his company had carried out.

Thanks to Stefan Waldhauser from BECA, Michelle Lucas from Te Kura, and Nicole Humphries from Waitakere Properties for allowing us to carry out user surveys of their buildings, Aorangi House, and Waitakere Central One. Also to all of the building occupants that took the time to fill in the questionnaire, as without their willing responses this study would not be possible.

Finally I would like to thank all my family and friends for their continuous support throughout this research.

TABLE OF CONTENTS

Preface	I
Abstract	III
Acknowledgments	V
List of Tables, Figures, & Equations	XI
List of Tables	XI
List of Figures	XIII
List of Equations	XVII
Abbreviations.....	XIX
Definitions	XXI
1.0 Introduction	1
1.1 Background/Issues	1
1.2 Aim	2
1.3 Hypothesis	2
1.4 Objectives.....	2
1.5 Outline methodology of the research	2
1.6 Outline of thesis	3
2.0 Background	5
2.1 How the building environment affects the health, comfort, and productivity of building occupants5	
2.1.1 Health.....	7
2.1.2 Comfort	8
2.1.3 Productivity	9
2.2 Methods of assessing the indoor environmental quality	10
2.2.1 Questionnaires as a tool	11
2.2.2 Categories	12
2.2.3 Rating scales.....	13
2.3 Examples of questionnaires used	15
2.4 Background summary.....	19
3.0 Benchmarks	21
3.1 Nature and characteristics of benchmarks	21
3.1.1 General statistics.....	22
3.2 Benchmarks already applied to the building industry.....	25
3.2.1 Energy benchmarks.....	26
3.3 Benchmarks of people’s perceptions	28
3.3.1 Benchmarks from examples of questionnaires used	30
3.4 Assessment of BUS benchmarks	35
3.4.1 Indices	38
3.4.2 BUS rating scores	40
3.5 Current building rating tools	40
3.5.1 Existing building rating tools	41
3.5.2 Comparison of the indoor environmental quality category	44
3.5.2 Current building rating tools that incorporate perceptions	46
3.6 Benchmark summary	49
4.0 Research Methods	51
4.1 BUS questionnaire design	51
4.2 Pilot database.....	53
4.2.1 Previously surveyed buildings.....	53
4.2.2 Buildings surveyed during this study	55
4.2.3 Buildings that could be surveyed in the future - sourced from BEES project	55

4.3 Analysis of scores	56
4.3.1 Classification of design	56
4.3.2 Benchmark comparison	57
4.3.3 Questionnaire scores	58
4.3.4 Questionnaire variables	59
4.3.5 Building variables	60
4.4 Exploring benchmark format and incorporation into BSRTs	62
5.0 Analysis of Scores	63
5.1 Classification of design	63
5.1.1 New Zealand building stock	63
5.1.2 Pilot database	64
5.2 Comparison of benchmarks and databases	69
5.2.1 How the pilot database's averages compare to the current BUS benchmark being used	69
5.2.2 Comparison of buildings in terms of current BUS benchmark and dataset	71
5.2.3 Summary	74
5.3 Overview of scores of the buildings	75
5.3.1 Distribution	75
5.3.2 Skewness	92
5.3.3 Kurtosis	93
5.3.4 Normal distribution test	94
5.3.5 Summary	95
5.4 Questionnaire variables	96
5.4.1 Correlations	96
5.4.2 BUS summary method comparison	98
5.5 Building variables	103
5.6 Analysis of scores summary	106
6.0 User perception benchmarks and their incorporation into rating tools	109
6.1 Questionnaire variables	110
6.2 Building 21 and Building 29 ratings with BUS benchmarks	110
6.3 Absolute benchmarks	111
6.3.1 Percentage of people satisfied/dissatisfied	111
6.3.2 Individual Scores	113
6.3.3 Scores on the 7-point scale	114
6.3.4 Summary of absolute benchmarks	117
6.4 Relative Benchmarks	118
6.4.1 Distribution of scores in relation to the pilot database	118
6.5 Combined benchmarks	120
6.5.1 In relation to the scales and the pilot database	120
6.5.2 Alternate way of incorporating both the scale and the benchmark	121
6.5.3 Comfort Index	124
6.5.4 Summary of relative and combined benchmarks	125
6.6 User perception benchmarks summary	126
6.7 Incorporation into rating tools	127
7.0 Conclusions and Recommendations	131
7.1 Conclusions	131
7.1.1 Determining users' perceptions	131
7.1.2 Benchmarks	132
7.1.3 Pilot database	133
7.1.4 Exploring benchmarks and incorporation methods	133

7.2 Other key factors.....	134
7.3 Future work	134
Works Cited	135
Appendices	143
Appendix A: Questionnaire based methods.....	143
Appendix B: BUS rating score method	145
Appendix C: Adequate sample size calculation	146
Appendix D: Normality tests	148
Appendix E: Correlations	149
Appendix E.1: Individual variables correlations	149
Appendix E.2: Correlations of variables to indices.....	150
Appendix E.3: Correlations to BUS rating score: selected variable method	150
Appendix F: Building variables	151
Appendix F.1: Independent t-tests	151
Appendix G: Individual building distributions	154
Appendix G.1: Building 21 – low scoring building.....	154
Appendix G.2: Building 29 – high scoring building.....	155
Appendix H: Internal environment variables – distributions	157
Appendix H.1: Conditions in winter – <i>Temperature</i> and <i>Air Quality</i>	157
Appendix H.2: Conditions in summer – <i>Temperature</i> and <i>Air Quality</i>	158
Appendix H.3: Noise.....	159
Appendix H.4: Lighting	160
Appendix H.5: Space	161

LIST OF TABLES, FIGURES, & EQUATIONS

List of Tables

Table 1: Summary of common scales (Oseland, 2007, p. 32)	14
Table 2: Questionnaire examples.....	16
Table 3: Questionnaire based methods' benchmarks.....	29
Table 4: Green Star rating tool scores	42
Table 5: Rating tool workshop environmental aspect weighting (NZGBC, 2006, p. 25)	45
Table 6: NABERS protocol (Baird, 2009, p. 1073)	46
Table 7: HOPE classifications (Roulet, Flourentzou, Foradini, Bluyssen, Cox, & Aizlewood, 2006)	48
Table 8: POE details of the previously surveyed buildings	54
Table 9: POE details of the buildings surveyed during this study.....	55
Table 10: Classification of design categories	57
Table 11: Size effects (Salkind, 2005, p. 129)	60
Table 12: Example of building classification: TOBUS methodology (Caccavelli & Gugerli, 2002, p. 6).....	66
Table 13: Classification of design features	67
Table 14: Sustainable buildings' design features and awards	68
Table 15: Descriptive statistics of Design scores.....	76
Table 16: Descriptive statistics of Needs scores.....	77
Table 17: Descriptive statistics of Space scores	78
Table 18: Descriptive statistics of Image scores	79
Table 19: Descriptive statistics of Meeting Rooms scores	80
Table 20: Descriptive statistics of Storage scores	81
Table 21: Descriptive statistics of Furniture scores.....	82
Table 22: Descriptive statistics of Temperature Overall in Winter scores	83
Table 23: Descriptive statistics of Temperature Overall in Summer scores	84
Table 24: Descriptive statistics of Air Overall in Winter scores.....	85
Table 25: Descriptive statistics for Air Overall in Summer scores	86
Table 26: Descriptive statistics of Noise Overall scores	87
Table 27: Descriptive statistics of Lighting Overall scores	88
Table 28: Descriptive statistics of Overall Comfort scores	89
Table 29: Descriptive statistics of Perceived Productivity scores.....	90
Table 30: Descriptive statistics of Health scores.....	91
Table 31: Normality test in terms of skewness	92
Table 32: Normality test in terms of kurtosis.....	93
Table 33: Kolmogorov-Smirnov test of normality	94
Table 34: Correlations between summary variables.....	97
Table 35: Independent t-test in relation to sustainability.....	104
Table 36: Building features	109
Table 37: Current benchmark scores	111
Table 38: Individual scores method	113
Table 39: Bands of the scale	115
Table 40: Benchmarks from absolute scale method.....	116
Table 41: Benchmarks for quartile/quintile method.....	119
Table 42: Benchmarks for combined scale and percentile method	120
Table 43: Building 21 - BUS rating score: selected variable method	121
Table 44: Building 21 – BUS rating score: all variable method	122
Table 45: Building 29 – BUS rating score: all variable method	123
Table 46: Splits in Comfort Index scale	125

<i>Table 47: Incorporation example using percentage satisfied</i>	<i>127</i>
<i>Table 48: NABERS star ratings</i>	<i>128</i>
<i>Table 49: Incorporation example using the altered BUS rating score: all variable method</i>	<i>129</i>
<i>Table 50: Examples of questionnaire based methods.....</i>	<i>143</i>
<i>Table 51: References for questionnaire examples</i>	<i>144</i>
<i>Table 52: Kolmogorov-Smirnov test results</i>	<i>148</i>
<i>Table 53: Correlation of summary variables.....</i>	<i>149</i>
<i>Table 54: Correlations of indices and summary variables</i>	<i>150</i>
<i>Table 55: Correlations of BUS rating score: selected variable method and summary variables</i>	<i>150</i>
<i>Table 56: Independent samples t-test group statistics</i>	<i>151</i>
<i>Table 57: Independent samples t-test results.....</i>	<i>152</i>

List of Figures

Figure 1: 20 year present values for an office building (Fullbrook, Jackson, & Finlay, 2006, p. 45).....	5
Figure 2: Sample size required for 5% margin of error and 95% confidence level (Oseland, 2007, p. 28).....	11
Figure 3: Types of scales in Oseland's sample.....	18
Figure 4: Types of scales in 20 questionnaires studied	18
Figure 5: Example semantic differential scale (Center for the Built Environment, 2009)	18
Figure 6: Example labelled/categorical rating scale (Swanke Hayden Connell Architects, 2005)	18
Figure 7: Normal distribution example (Aronoff & Kaplan, 1995, p. 353)	23
Figure 8: Standard deviation example (Peck, Olsen, & Devore, 2009, p. 180).....	24
Figure 9: Quartile example (Mendenhall, Beaver, & Beaver, 2009, p. 77)	25
Figure 10: Labelling scale and cumulative frequency curve (Perez-Lombard et al. 2009, p.276)	28
Figure 11: Example of radial diagram (Swanke Hayden Connell Architects, 2005).....	33
Figure 12: OLS fingerprint of performance (ABS Consulting, 2008, p. 4)	34
Figure 13: Ranking of individual buildings (Alexi Marmot Associates, 2004)	34
Figure 14: Example of bar chart (Oseland, 2007)	35
Figure 15: Distribution of scores (Building Use Studies, 2009)	36
Figure 16: Critical region (Building Use Studies, 2009)	37
Figure 17: Percentile graph (Building Use Studies, 2009)	38
Figure 18: Basic structure of Green Star NZ (NZGBC, 2008)	42
Figure 19: New Construction (USGBC, 2011)	43
Figure 20: Existing Buildings (USGBC, 2011).....	43
Figure 21: CASBEE example (Boonstra & Pettersen, 2003, p. 82).....	44
Figure 22: Comparison of environmental aspects weightings (Smith, 2008, p. 43).....	45
Figure 23: Noise section of POE questionnaire	52
Figure 24: Skewness (Salkind, 2005, p. 97)	59
Figure 25: Kurtosis (Salkind, 2005, p. 98).....	59
Figure 26: Distribution of building types by floor area (Isaacs, 2008, p. 45)	63
Figure 27: Distribution of building types by numbers (Isaacs, 2008, p. 45)	64
Figure 28: Proportions of pilot buildings around NZ.....	64
Figure 29: Comparison between NZ building stock and pilot database in terms of size	65
Figure 30: Ventilation distribution	69
Figure 31: Comparison of databases according to summary variables	69
Figure 32: Comparison of databases including all questions.....	71
Figure 33: Summary variable comparison to benchmarks regardless of time of study	72
Figure 34: Summary variable comparison to UK 2006 benchmark.....	73
Figure 35: Summary variable comparison to NZ 2010 benchmark	74
Figure 36: Display example	75
Figure 37: Frequency distribution of mean Design scores	76
Figure 38: Frequency distribution of mean Needs scores	77
Figure 39: Frequency distribution of mean Space scores.....	78
Figure 40: Frequency distribution of mean Image scores	79
Figure 41: Frequency distribution of mean Meeting Rooms scores.....	80
Figure 42: Frequency distribution of mean Storage scores.....	81
Figure 43: Frequency distribution of mean Furniture scores	82
Figure 44: Frequency distribution of mean Temperature Overall in Winter scores	83
Figure 45: Frequency distribution of mean Temperature Overall in Summer scores.....	84
Figure 46: Frequency distribution of mean Air Overall in Winter scores	85
Figure 47: Frequency distribution of mean Air Overall in Summer scores	86
Figure 48: Frequency distribution of mean Noise Overall scores.....	87

Figure 49: Frequency distribution of mean Lighting Overall scores.....	88
Figure 50: Frequency distribution of mean Overall Comfort scores.....	89
Figure 51: Frequency distribution of mean Perceived Productivity scores.....	90
Figure 52: Frequency distribution of mean Health scores	91
Figure 53: Smallest level of skewness	92
Figure 54: Highest level of skewness	92
Figure 55: Negative kurtosis	94
Figure 56: Positive kurtosis	94
Figure 57: Highest correlation in sample - Design and Image	96
Figure 58: UK Comfort Index for pilot database buildings.....	99
Figure 59: UK Satisfaction Index for pilot database buildings	99
Figure 60: UK Summary Index for pilot database buildings.....	99
Figure 61: UK Forgiveness Factor of pilot database buildings	99
Figure 62: NZ Comfort Index for pilot database buildings	100
Figure 63: NZ Satisfaction Index of pilot database buildings.....	100
Figure 64: NZ Summary Index for pilot database buildings.....	100
Figure 65: NZ Forgiveness Factor for pilot database buildings.....	100
Figure 66: BUS rating score: selected variable method results for pilot database buildings when compared to BUS benchmarks	102
Figure 67: Large difference - Image	104
Figure 68: Large difference - Perceived Productivity	104
Figure 69: No significant difference - Storage	105
Figure 70: No significant difference - Noise Overall.....	105
Figure 71: Percentage of people satisfied/dissatisfied - Building 21	112
Figure 72: Percentage of people satisfied/dissatisfied - Building 29	112
Figure 73: HOPE	115
Figure 74: Quarters.....	115
Figure 75: Fifths	115
Figure 76: Sixths.....	115
Figure 77: Quartiles	118
Figure 78: Quintiles.....	118
Figure 79: Comfort Index quartiles	125
Figure 80: Comfort Index quintiles.....	125
Figure 81: Comparison of different benchmarks for Absolute, Relative and Combined methods	126
Figure 82: Example of a 5 rating.....	145
Figure 83: Example of a 1 rating.....	145
Figure 84: Example of a 3 rating.....	145
Figure 85: Example of a 4 rating.....	145
Figure 86: Example of a 2 rating.....	145
Figure 87: Building 21 - Temperature Overall in Winter	154
Figure 88: Building 21 - Temperature Overall in Summer.....	154
Figure 89: Building 21 - Air Overall in Winter	154
Figure 90: Building 21 - Air Overall in Summer	154
Figure 91: Building 21 - Lighting Overall.....	154
Figure 92: Building 21 - Noise Overall.....	154
Figure 93: Building 21 - Overall Comfort	155
Figure 94: Building 21 - Space.....	155
Figure 95: Building 29 - Temperature Overall in Winter	155
Figure 96: Building 29 - Temperature Overall in Summer.....	155

<i>Figure 97: Building 29 - Air Overall in Winter</i>	<i>155</i>
<i>Figure 98: Building 29 - Air Overall in Summer</i>	<i>155</i>
<i>Figure 99: Building 29 - Overall Comfort</i>	<i>156</i>
<i>Figure 100: Building 29 - Space.....</i>	<i>156</i>
<i>Figure 101: Temperature in Winter: Hot/Cold.....</i>	<i>157</i>
<i>Figure 102: Temperature in Winter: Stable/Varies.....</i>	<i>157</i>
<i>Figure 103: Air in Winter: Still/Draughty</i>	<i>157</i>
<i>Figure 104: Air in Winter: Dry/Humid.....</i>	<i>157</i>
<i>Figure 105: Air in Winter: Fresh/Stuffy</i>	<i>157</i>
<i>Figure 106: Air in Winter: Odourless/Smelly.....</i>	<i>157</i>
<i>Figure 107: Temperature in Summer: Hot/Cold.....</i>	<i>158</i>
<i>Figure 108: Temperature in Summer: Stable/Varies</i>	<i>158</i>
<i>Figure 109: Air in Summer: Still/Draughty.....</i>	<i>158</i>
<i>Figure 110: Air in Summer: Dry/Humid.....</i>	<i>158</i>
<i>Figure 111: Air in Summer: Fresh/Stuffy.....</i>	<i>158</i>
<i>Figure 112: Air in Summer: Odourless/Smelly</i>	<i>158</i>
<i>Figure 113: Noise from Colleagues</i>	<i>159</i>
<i>Figure 114: Noise from Others.....</i>	<i>159</i>
<i>Figure 115: Other Noise from Inside</i>	<i>159</i>
<i>Figure 116: Noise from Outside</i>	<i>159</i>
<i>Figure 117: Unwanted Interruptions</i>	<i>159</i>
<i>Figure 118: Natural Light Conditions.....</i>	<i>160</i>
<i>Figure 119: Glare from Sun and Sky.....</i>	<i>160</i>
<i>Figure 120: Quality of Artificial Light.....</i>	<i>160</i>
<i>Figure 121: Glare from the Lights.....</i>	<i>160</i>
<i>Figure 122: Space at Desk.....</i>	<i>161</i>

List of Equations

Equation 1: Standard score.....	24
Equation 2: TOBUS comfort index (Bluyssen & Cox, 2002)	33
Equation 3: Standard score for BUS data	38
Equation 4: BUS Comfort Index (Building Use Studies, 2008)	38
Equation 5: BUS Satisfaction Index (Building Use Studies, 2008)	39
Equation 6: BUS Summary Index (Building Use Studies, 2008)	39
Equation 7: BUS Forgiveness Factor (Building Use Studies, 2008)	39
Equation 8: Calculating effect size (Pallant, 2005)	61
Equation 9: Comfort Index for Building 21 using pilot database	124
Equation 10: Comfort Index for Building 29 using pilot database	124
Equation 11: Sample size calculation.....	146

ABBREVIATIONS

AC – Air Conditioning
BSI – Building Symptom Index
BSRT – Building Sustainability Rating Tool
EIA – Energy Information Administration
EPI – Energy Performance Index
EUI – Energy Use Index
IAQ – Indoor Air Quality
IEQ – Indoor Environmental Quality
K-S – Kolmogorov-Smirnov
MM – Mixed Mode
NV – Natural Ventilation
PMV – Predicted Mean Vote
POE – Post Occupancy Evaluation
PPD – Percentage of People Dissatisfied
PSI – Personal Symptom Index
SBS – Sick Building Syndrome
SD – Standard Deviation
SPSS – Statistical Package for the Social Sciences

Organisations

AMA – Alexi Marmot Associates
ASHRAE – American Society of Heating, Refrigeration and Air-Conditioning Engineers
BRE – Building Research Establishment
BUS – Building Use Studies
CBE – Center for the Built Environment
CIBSE – Chartered Institute of Building Services Engineers
EECA – Energy Efficiency and Conservation Authority
IPD – Investment Property Databank
NZGBC – New Zealand Green Building Council
USBGC – United States Green Building Council
VUW SoA – Victoria University of Wellington, School of Architecture

Other studies

BEES – Building Energy End-Use Study
BREEAM – Building Research Establishment Environmental Assessment Method
CASBEE – Comprehensive Assessment System for Built Environment Efficiency
COPE – Cost-Effective Open Plan Environments
DQI – Design Quality Indicator
HOPE – Health Optimisation Protocol for Energy-efficient Buildings
LEED – Leadership in Energy and Environmental Design
NABERS – National Australian Built Environment Reporting System
OLS – Overall Liking Score
OPN – Office Productivity Network
PROBE – Post-Occupancy Review of Buildings and their Engineering
SCATS – Smarts Control and Thermal Comfort Study
TOBUS – Tool for Selecting Office Building Upgrading Solutions

DEFINITIONS

Confidence Interval – a range of values defined so that there is a specified probability that the value of a parameter lies within it.

Confidence Limits – the lower and upper boundaries of a confidence interval; that is the values which define the range of a confidence interval.

Confidence Level - the probability that the value of a parameter falls within a specified range of values.

Eta Squared – represents the proportion of variance that is explained by the independent variable.

Homogeneity of Variance - looks at the variance between the different groups and whether or not it is equal.

Independent t-test – procedure for determining whether significant differences exist between two sample means.

Kolmogorov-Smirnov test – procedure for determining whether a distribution can be considered to approximate a normal distribution. It 'takes the observed cumulative distribution of scores and compares them to the theoretical cumulative distribution for a normally distributed population' (Hinton, Brownlow, McMurray, & Cozens, 2004, p. 30)

Kurtosis - information about 'peakedness' of distribution

Mean – measure of central tendency, average of the scores

Pearson correlation coefficient – a numerical summary of the strength and direction of the relationship between two variables.

Skewness - indication of the symmetry of distribution, one tail is longer than the other.

Standard Deviation – a measure of variability which represents the degree in which the data is spread around the mean, and represents the average amount of variability in a set of scores.

Z-score – a standard score which measures how many standard deviations the value is from the mean.

1.0 INTRODUCTION

‘To be a little blunt, and perhaps obvious, buildings are not put up in order to be sustainable or to reduce energy consumption, they are built and have value if they support activities that people want and need to carry out’ (Hyams, 2004).

This introductory chapter discusses the background to the research project, and presents its aim and objectives, before giving a brief outline of the methodology and finally an outline of the thesis.

1.1 Background/Issues

Buildings are ultimately built for people to use, so it is important that they are designed to suit the occupants’ needs. According to the NZ Medical Journal (2007), New Zealanders spend a total of 94% of time indoors in buildings or vehicles (Baker, Keall, Au, & Howden-Chapman, 2007); hence the internal environment will have a significant impact on its occupants. It is therefore desirable that all buildings provide healthy, comfortable environments.

Consideration of building performance is becoming more mainstream in the building industry, largely due to the recent trend towards buildings that are ‘sustainable.’ There are numerous definitions of the term ‘sustainable’, but generally this type of design aims to provide a good level of interior environment while having minimal effect on the environment. A high level in interior quality is important ‘because buildings are social spaces driven by the needs of organisations that work in them. If a ‘green’ building does not provide the functionality that the users need, or looks out of place with its context, it will either be removed, or significantly renovated, which is not sustainable’ (Sustainable Built Environments, Centre for Design at RMIT University, 2007). Building Sustainability Rating Tools (BSRTs) are increasingly used to demonstrate how the buildings achieve their sustainable aim.

Predominantly, current BSRTs have their main focus on technical aspects of the building such as energy consumption and materials recycling; very little attention is paid to how the occupants themselves rate the building. User perception benchmarks will allow building owners/users to see how New Zealand buildings are rated by the occupants themselves. Although subjective, it will allow the rating of what the space is actually designed for, the occupants.

‘Gary Raw, who ran the Healthy Building Centre at the Building Research Establishment (BRE), has said: “People are the most valid measuring instruments: they are just harder to calibrate”. With careful sampling even seemingly contradictory responses from wide ranges of people can yield useful information’ (Leaman, 2004, p. 503).

The assessment of performance of buildings is not limited to ‘sustainable’ buildings, but can be applied to any existing building. Only a small percentage of the working population gets to move into a building that is newly constructed. Therefore the existing building stock is the most appropriate to assess, as it will affect the largest number of people. However, at present there is no statistically unbiased, country-based benchmark for the perceptions of the occupants of the building environments that is representative of the existing building stock in New Zealand.

‘Obviously user sensibility and satisfaction must play a pre-eminent role in evaluating all types of facilities and therefore they must play an active part in building performance of all types’ (Meir, Garb, Jiao, & Cicelsky, 2009, p. 208).

The literature shows a considerable amount of research effort has been directed at the performance of buildings from the point of view of the occupant and ultimate end-users of the buildings, however it is still is not in the forefront of building evaluation, despite being fundamental to the overall success of a building environment.

1.2 Aim

The aim of this study is to assist in the improvement of the performance of commercial and institutional buildings for their users through the development of user perception benchmarks.

1.3 Hypothesis

Given a representative sample of building user questionnaires it should be feasible to develop user perception benchmarks for New Zealand buildings and incorporate them into building rating tools.

1.4 Objectives

- To provide a general overview of existing studies which have investigated user perceptions.
- To explore the nature and characteristics of benchmarks in general; those in which people's perceptions play a key part, and those already applied in the building industry.
- To create a pilot database of results from user questionnaires of New Zealand Buildings.
- To use the data collected to explore methods for determining benchmarks for NZ commercial and institutional buildings in terms of user perception.

The research endeavours to analyse the data received from a sample of surveyed New Zealand buildings to determine patterns and influencing factors. This research is also intended to be added to in future work to eventually provide a profile of the existing New Zealand commercial and institutional building stock, establishing a base point for future measurement.

1.5 Outline methodology of the research

Post Occupancy Evaluations (POEs) allow some insight into how the occupants perceive conditions in a building and any problems that they encounter. The responses are subjective, as everyone has differing ideas on comfort, but common trends often allow problem areas to be identified and a general rating of the building to be determined. The POE data that will be used in this project is from studies carried out using the Building Use Studies (BUS) Methodology (Building Use Studies, 2009) which elicits responses on 45 aspects of building performance.

A pilot database of results from the surveyed New Zealand buildings was compiled. This allowed flexibility in the statistical analysis, and exploration of how possible benchmarks could be constructed. From this database the questionnaire data received was explored as to the type of data that the questionnaire gave back, and relationships between responses to questions investigated. The data was explored to see whether the design of the buildings had a significant impact on user perceptions. These analyses enabled a discussion on how the data could be made into benchmarks and how these could be incorporated into building rating tools.

The original intention was to create a profile of the commercial and institutional buildings in New Zealand, which were being reviewed as part of a separate major research project (BEES), and create some statistically sound benchmarks from this representative sample. As the data was not available at the time this study was carried out, the project focus was shifted to an exploration of the data available from a smaller sample of buildings.

1.6 Outline of thesis

This research report is divided into seven main sections, with the current **Chapter 1.0: Introduction** introducing the topic and issues, as well as providing the research aim, hypothesis and objectives.

Chapter 2.0: Background gives a brief outline of the internal built environment and how it affects occupants, and explores the methodology of surveying, including an analysis of some previously established methods which have been applied to the building industry.

Chapter 3.0: Benchmarks explores the nature and characteristics of benchmarks in general, those already applied in the building industry, those in which people's perceptions play a key part, and an assessment of the currently used Building Use Studies (BUS) benchmarks.

The methodology that is employed in this project is presented in **Chapter 4.0: Research methods**. It includes a description of the questionnaire, the buildings that form the pilot database and the procedures for completing the analysis.

Chapter 5.0: Analysis of scores examines the scores received from the surveying of the pilot database. Firstly, a comparison of the current benchmark databases is made, followed by further analysis of the distribution of scores. Lastly, the effect of the building design on the resultant scores is briefly considered.

Chapter 6.0: User perception benchmarks and their incorporation into rating tools investigates potential approaches for the creation of benchmarks for users' perception scores and the incorporation of these scores into rating tools.

Chapter 7.0: Conclusions and recommendations summarises the findings from this analysis to draw conclusions and make further recommendations into how this area of indoor environmental quality can be dealt with, and possible ways of incorporating user perception benchmarks into building rating tools.

2.0 BACKGROUND

Chapter 2 discusses the built environment's effect on the occupants, ways to evaluate the building environment and examples of previously used evaluation questionnaires. The chapter is divided into four sections:

- Section 2.1 investigates how occupants can be affected by the built environment, with particular focus on working environments and the effect on comfort, health and productivity.
- Section 2.2 explores methods of assessing the indoor environment, with a particular focus on the use of questionnaires.
- Section 2.3 investigates previously used questionnaire based methodologies for the assessment of the built environment.
- Section 2.4 provides a summary discussion of the literature, with a focus on the supporting evidence of the approach taken in this research.

2.1 How the building environment affects the health, comfort, and productivity of building occupants

'In developed countries people spend more than 90% of their time indoors. Indoor conditions have therefore far-reaching implications for their health, general well-being and performance' (Frontczak & Wargocki, 2010, p. 1).

People spend a lot of time in buildings. In particular, it is not uncommon for people to spend 8 hours of the day at work in offices and institutional buildings; therefore it is imperative that the buildings are not having a negative effect on them, and are conducive to a good working environment.

'The better the quality of the physical work setting, the less aware people are of its contribution. At its best, the office environment is a seamless background that the occupants hardly notice. At its worst, it is a major obstacle to productive effort – a place where no real work can be done' (Aronoff & Kaplan, 1995, p. 1).

In conjunction with the great amount of time occupants spend in workplaces, the importance of a good indoor environment can also be brought into the commercial setting. Although making a building energy efficient has direct financial gain that you are able to see, it is not in fact the biggest cost in a commercial building. Salaries of the employees far outweigh any of the other costs of the building. Fullbrook et al (2006) considered the relative present values for the different costs associated with a business, and found that salary cost was over three-quarters of the total. The results from their analysis are shown in the pie chart in Figure 1 below.

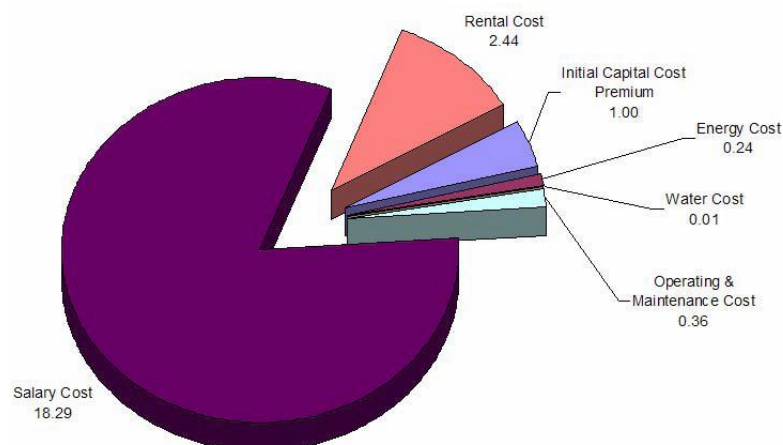


Figure 1: 20 year present values for an office building (Fullbrook, Jackson, & Finlay, 2006, p. 45)

The fact that the salary cost is the largest, should result in employers wanting their employees to work to their potential; one way to do this is to provide a healthy, comfortable environment that does not have an adverse effect on them, and which may in fact improve their productivity.

The indoor environment in a building will be largely affected by the country it is in. The climate for which a building is designed should have a direct influence on the type of designs that are produced. For example, in really hot, humid countries air conditioning or a more complex form of natural ventilation needs to be used as the direct transfer of air from outside to inside will not provide a comfortable temperature for the occupants. Cultural expectations, even dress codes will also affect the user perceptions of the internal environment.

'Inevitably, the question arises about global similarities and differences, especially in building users' attitudes and preferences, but also in comparisons between buildings themselves. For example, British buildings seem to be more 'stressed', with higher occupation densities, more likelihood of open plan layouts and an increasing tendency to 24hour/7day operation. Features in offices which occupants like (such as lower densities, cellularisation, natural light and controllability) seem to be less common than they are in mainland Europe. This might mean that British buildings overall compare unfavourably, for instance, with their European counterparts. However, British buildings, at least from their occupants' perspective, seem to be improving, albeit slowly' (Leaman, 2003).

There has been a vast amount of research conducted into the effect of the indoor environmental quality on occupants. The indoor environment of a building can affect the occupants in different ways. The three aspects that are investigated further in this report are; Health, Comfort, and Productivity, due to the great affect they can have on the occupant. These are often quite difficult to quantify and it is hard to isolate problematic aspects as they are all affected by many different variables to do with the environmental features, ergonomics, design, as well as differences in social and psychological aspects.

'There are four basic environmental factors in the indoor environment that directly influence the perception of that indoor environment through the senses, but also have an effect on the physical and mental state (comfort and health) of occupants:

1. **Thermal comfort or indoor climate**, comprising parameters such as moisture, air velocity and temperature
2. **Visual or lighting quality**, determined by view, illuminance, luminance ratios, reflection and other parameters
3. **Indoor air quality**: a complex phenomenon comprising odour, indoor air pollution, fresh air supply, etc.
4. **Acoustical quality**, influenced by outside and inside noise, as well as vibrations.

In addition, ergonomics, such as the dimensions and sizes of the space, tools, furniture, etc., play an important role in total body perception' (Bluyssen P. , 2009, p. 45).

Job stress, management and social factors will also affect the satisfaction of occupants; however as this study is focussed on the building itself, the four basic environmental factors are the main focus of **Section 2.1: How the building environment affects the health, comfort, and productivity of building occupants.**

2.1.1 Health

When people are spending long periods of time in one building or space, that indoor environment can affect health. 'Studies suggest that 20% to 30% of commercial buildings have problems that lead to occupant complaints and illness' (Aronoff & Kaplan, 1995). This could be due to many different factors, including environmental systems or the materials used in the space. Increased chemical pollutants in the air or from surfaces can cause health problems, while some of the most common problems are caused by the air quality in the building. Often there are factors that can be easily improved.

In the past, there have been many issues with buildings affecting the health of its occupants.

'Sick building syndrome (SBS), sometimes called tight building syndrome, is a condition in which occupants experience an undue number of illness symptoms when inside a building that diminish when they leave' (Aronoff & Kaplan, 1995).

Common symptoms of this include headache, fatigue, drowsiness, and nose, throat, and eye irritation. Many physical factors such as air quality, relative humidity, temperature, noise, vibration, and lighting have been linked with sick buildings. The most commonly linked with sick buildings are poor indoor air quality and inadequate ventilation.

- **Thermal Conditions:** The effect of the thermal environment on the occupant is sometimes framed as 'thermal stress.' Although they are not common occurrences in the commercial environment, some are still able to have an effect. Exposure to hot conditions can cause dehydration, heat stress and eventually heat stroke, while exposure to cold conditions can affect dexterity and mobility and may increase physical and visual strain. Low levels of humidity can exacerbate respiratory and skin conditions (London Hazards Centre, 1995).
- **Lighting Quality:** There are several lighting variables that can affect the health of an occupant in terms of both natural and artificial. Spending long periods of time under fluorescent lighting is associated with headaches, eye-strain, eye-irritation, fatigue and increased stress and accidents. With natural lighting, it is mainly the shortage of natural light, or too much in the form of glare, that becomes a problem (London Hazards Centre, 1995).
- **Indoor Air Quality:** This is most commonly considered to be a cause of SBS. Indoor air quality deals with pollutants in the air. In the commercial setting, due to the long periods of time people spend in them, even seemingly low levels can affect people's health. Inadequate ventilation and excess moisture contribute to asthma and mould induced illnesses.
- **Acoustical Quality:** Noise is not such a high level of concern in regards to health in buildings, as the noise produced in offices will not very often get to levels of concern. It can however have an effect on the health of occupants. Excessive noise exposure can cause hearing loss. The extent of this depends on the intensity of noise, its duration and its frequency (NZ Department of Labour, 2009).

Occupants are also pollutants of the space. They give off heat and CO₂ which will increase the temperature and affect the air quality. The environmental systems need to be catering for the right amount of people, as more people in the space will require more fresh air. The ventilation system also needs to take into account the different seasons of the year.

'Most health problems associated with the office environment result from chronic exposure to a low-level stressor rather than sudden exposure to an acute hazard' (Aronoff & Kaplan, 1995, p. 157).

2.1.2 Comfort

Comfort, similar to health, in an indoor environment, is influenced by many variables. The temperature, air quality, noise, lighting, furniture all will affect the comfort of the occupants using the space. The different factors do not have to be at such an extreme level to cause people discomfort, to have a negative effect on their health.

'The list of variables may look daunting and the chance of satisfying everyone is slim. The accepted measure of success is where 80% of the occupants are thought to be comfortable' (Dwyer, 2006, p. 90).

- **Thermal Comfort:** Thermal comfort is particularly subjective. It is influenced by a range of factors such as metabolic rate, clothing, air temperature, radiant temperature of surrounding surfaces, rate of air flow etc. Indices have been created to take these into account. There are also broadly accepted 'comfort temperatures', such as those determined by the World Health Organisation (World Health Organisation, 2011).
- **Lighting Quality:** This can also be termed 'visual comfort'. A main feature that is considered to be conducive of a comfortable environment is the use of adequate natural lighting, as this provides a view to the outside. Artificial lighting can also affect comfort; this could be through contrasting light levels, distracting flickering, colour rendering, humming. Glare, at lower levels than causing damage to health, can also be a cause for discomfort.
- **Indoor Air Quality:** Comfort in terms of indoor air quality is often the level of fresh air and air velocity and movement. This can be to do with whether it is too draughty or too stuffy. Odours in the indoor air quality can also affect the comfort of the occupant.
- **Acoustical Quality:** Although noise in office buildings is not often a cause for concern in terms of health, it is one of the greatest sources of complaints. This can be due to the two main issues of distraction and privacy. Distracting noise is not only dependent on the volume and the duration, but also the information carried by the noise. Privacy generally describes the ability to talk in one space without being overheard in another (CIBSE, 2005; Oesterle, Rolf-Pieter, Lutz, & Heusler, 2001).

A major factor that has been determined in occupant comfort in buildings is the amount of control that the occupants have. It has been shown that if the occupants feel they have control of some aspects they will be more likely be satisfied. 'The privileged few, who have much more control over both their environmental systems and their use of time and space in the building, tend to be happier' (Leaman & Bordass, 2007, p. 665). This could be as simple as having openable windows, so when an occupant gets hot or stuffy he or she is able to open a window straight away without waiting for a change in the air conditioning. However, control is less of a concern when things are working well in buildings; for example there are a number of buildings where users have low levels of control, but rate overall comfort high (Leaman, Thomas, & Vandenberg, 2007).

Occupants are generally more concerned with not being uncomfortable, rather than being comfortable with a particular environment.

'It has been found that it often makes more sense for designers to try and remove sources of user dissatisfaction rather than providing more things they hope users will like. The logic may sound a little strange but for building users the absence of, for example, thermal discomfort or unwanted interruptions is often more important. Building users are satisficers not optimizers, they adapt and cope' (Leaman & Bordass, 2007, p. 663).

2.1.3 Productivity

'Productivity, the highest potential gain of a healthy, comfortable, safe and secure space, has received a lot of attention in the past years.' (Bluyssen P. , 2009, p. 136).

Productivity is not completely dependent on the indoor environment as there are other factors, isolated from the indoor environment, which can have a great influence. However, the indoor environment has been shown to have quite an effect on it, and in particular the health and comfort of an occupant has a direct effect on their productivity. The impact the indoor environment performance has on the productivity of the occupants brings the issue into the commercial setting, as it has a direct financial impact.

'As opposed to sickness and accidents, fatigue and working reluctantly can affect a large number of people. This implies that the adverse effect on the performance of each individual does not have to be particularly substantial to result in a relatively high loss of productivity for the organisation as a whole' (Roelofsen, 2002, p. 248).

Productivity is quite difficult to measure but is often expressed in terms of efficiency, as the ratio of output to input, or the subjective rating of the occupants themselves. Leaman (2007) has shown through his user surveys of hundreds of buildings that there is a strong correlation between occupants' rating of overall comfort and their perceived productivity. This has also been shown in other studies where 'nearly two thirds of the occupants thought that a 10% or more increase in their productivity was possible by improving the office environment' (Clements-Croome & Baizhan, 2000, p. 633).

As productivity is directly influenced by health and comfort, the individual environmental factors also affect productivity. There have been numerous studies on individual factors of the indoor environment and their effect on productivity.

- **Thermal Comfort:**

'Many but not all studies indicate that small (few °C) differences in temperatures can influence workers' speed or accuracy by 2% to 20% in tasks such as typewriting, learning performance, reading speed, multiplication speed, and word memory' (Fisk, 2002, p. 59).

- **Lighting Quality:**

'There are many lighting variables that can affect performance. Several studies have shown that people conducting paper-based work and industrial work are more productive at high levels of illuminance (1000 lux) whereas VDU work requires lower levels (< 500 lux) in order to reduce glare, increase contrast and limit eye strain. For example, increasing the illuminance at paper based workers' desks from 500 to 1000 lux was shown to improve their performance by 2.8%' (Oseland, 1999, p. 2).

- **Indoor Air Quality:**

'A majority of the research studies indicate an average productivity loss of 10 per cent due to poor IAQ. Therefore by improving the IAQ, a conservative benefit of 6 per cent could readily be achieved' (Dorgan & Dorgan, 2006, p. 128).

- **Acoustical Quality:**

'In an ASID/Yankelovich study of 1,000 office workers conducted that year, 70 percent of the respondents said their productivity would increase if their offices were less noisy' (American Society of Interior Designers, 2005, p. 1).

2.2 Methods of assessing the indoor environmental quality

The previous section considered the great affect that the building environment can have on the occupant. This section discusses methods of assessing the indoor environmental quality, with particular focus on methods of determining how the occupants rate the building, and how they think it affects them. Assessing a building after it is occupied is often termed 'Post Occupancy Evaluation.'

There are many techniques that have been used to evaluate the performance of a building. Leaman et al (2010) note a few that they have found, in their experience, to work well:

- **'Expert walk-throughs, with expert discussions.** *These are quick and usually effective, but experts can be fooled. Inexperienced researchers may miss important things, and are also unable to provide immediate comments and feedback to their hosts, whose patience tends to run out more quickly. Learners therefore benefit by visiting sites with more experienced researchers, especially for the initial visit or where uncertainties arise.*
- **Measuring technical performance** of building fabric, services and systems. *Here it is important to concentrate on what really matters and not be side-tracked.*
- **Assessing environmental performance,** usually energy but increasingly water and indoor air quality.
- **Occupant survey questionnaires.** *People often miscast surveys of users as merely 'subjective', but, as Gary Raw said, 'People are the best measuring instruments. They are just harder to calibrate' (private conversation, 1995). The most important thing with occupant surveys, especially in smaller buildings, is to get a high response rate.*
- **Structured discussions interviews with participants.** *If needed at all, these are usually best when the results of occupant and other surveys are available and can form a basis for discussion and identify issues and pinch points. In our experience, focus groups that include a peer group of people can work well in non-domestic buildings. However, in housing, individual interviews are better, as focus groups can easily settle on certain gripes and be dominated by peer pressure' (Leaman, Stevenson, & Bordass, 2010, p. 570).*

Three leading building industry associations have developed a consensus document for standardised protocols for assessing building performance for energy, water and indoor environmental quality. They split it into three levels of intervention, each with increasing cost and accuracy. They suggest that the first level be applied to all buildings, the second to all buildings with high performance claims, such as 'sustainable' buildings, and the third mainly used for research studies. In terms of the assessment of indoor environmental quality, the document suggests the first level to be the questionnaire, as it is the easiest and least expensive step to evaluate IEQ (ASHRAE; USGBC; CIBSE, 2009).

The particular focus of this report is the questionnaire. The questionnaire deals with occupants' perceptions. According to the Oxford Dictionary, the two definitions of perception are 'the ability to see, hear, or become aware of something through the senses' and 'the way in which something is regarded, understood, or interpreted' (Oxford University Press, 2010). In terms of this research project, when referring to occupants' perceptions, it is more in line with the second definition; that is, what people think of their indoor environment.

In assessing their indoor environment the occupants will be affected by many factors apart from the building environment. If they like the employer then they may be more inclined to give a positive rating. The relationship between the occupant and the building is a complex one. They are not only affected by the management, their mood, and the indoor environment, but it has also been found that their perceptions change over time.

'Buildings and their occupants interact continually and the relationship between them is dynamic. By using occupants as part of the means of measuring buildings, those who conduct POEs should understand the changing nature of the relationship between people, the climate and buildings. Surveys are therefore measuring a moving target, and close comparisons based on such surveys need to take this into account' (Nicol & Roaf, 2005, p. 345).

2.2.1 Questionnaires as a tool

Questionnaires are useful tools as part of a survey methodology, when wanting to determine attitudes for a large population of people, and often reach a sample of as many as 2,000. They, like any research tool, have advantages and disadvantages.

'A carefully selected probability sample in combination with a standardised questionnaire offers the possibility of making refined descriptive assertions about a student body, a city, a nation, or any other large population' (Babbie, 2008, p. 303) .'

Questionnaires allow many questions on a subject, giving flexibility to the analysis. Having standardised questions, asked over a wide population, allows for comparisons between different groups. They are often quick and easy to complete so the respondent is more likely to fill them in.

'Surveys operate on the basis of statistical sampling; only extremely rarely are full population surveys possible, practical or desirable' (Fellows & Liu, 2003, p. 23). The sample needs to be representative. This is determined by the sample size and the response rate.

The higher the overall response rate, the less chance of a significant bias. The high response rate produces results that will be more representative of what is in the building. What is considered to be a good response rate is disputed in the literature, but it is generally agreed that the more important factor than having a high response rate, is having no response bias in the responses received. This risk can be reduced by setting up a proper sampling frame to ensure responses received are from all ranges of people.

'The number of observations needed for a sample to be considered "large" depends on the conventions of the discipline. In attitude surveys, a sample size of at least 50 is generally used' (Aronoff & Kaplan, 1995, p. 354).

There are general rules of thumb for sample size that depend on the size of the overall population and the level of precision (confidence level) that is aimed for. For a lot of questionnaire data the confidence level aimed for is 95%. The graph below shows the sample size needed for a given population to get a confidence level of 95% (Oseland, 2007, p. 28). A sampling procedure is then generally followed to ensure that the people selected to make up the sample cover the whole range to make it representative.

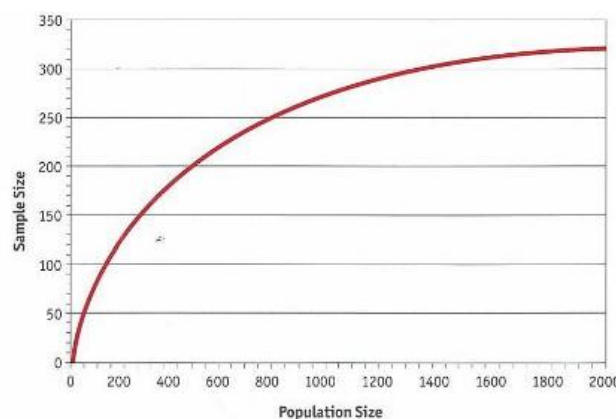


Figure 2: Sample size required for 5% margin of error and 95% confidence level (Oseland, 2007, p. 28)

It has been shown that one method of achieving a higher response rate is to give the respondents prior warning. This is even more effective when it comes from someone in a position of authority within the company, is confidential, and time during the working day is made for the completion of the questionnaire.

As the assessment is often carried out on workplaces, it is important that the questionnaires do not take up too much of the occupants' time, as it could be a factor in allowing the survey to be conducted in the first place.

'Experience suggests that for workplace evaluations a self-completion questionnaire which takes approximately 15 ± 5 minutes is acceptable' (Oseland, 2007, p. 30).

The questionnaires themselves can be made up of open or closed questions. The way a question is worded can determine the result given. Two questionnaires on the same topic may give different results and responses obtained may be very different due to a variety of factors.

'The way questions are phrased, the order in which they are asked, the format used to present them, and the response scales that are chosen can all bias the data that are collected' (Aronoff & Kaplan, 1995, p. 343).

The way each question is phrased can have an implied bias. For example, a question phrased 'how often is the noise in your office a problem', would get a much different response from the more neutral, 'how do you rate the noise levels in your office.' A questionnaire cannot appear to be steering the respondent in a certain direction as it will reduce its credibility as a valid measuring instrument.

The order in which the questions are placed can also have an effect on the responses. It is common practice to place the general questions first, and then follow up with the more detailed questions, as even if the respondent attempts to assess each question independently, a response to a question will be influenced by the questions that precede it (Aronoff & Kaplan, 1995).

Open questions are generally where the respondents are asked to reply by writing in response to a question or topic, where closed questions give the respondent specific answers to choose from. Closed questions allow for easier analysis of the results, while open questions cover all bases and reduce the rigidity of available responses, therefore lessening any constraints.

There are different types of closed questions. They are generally made up of either categories or rating scales.

'The type of response scales used in the questionnaire determines the type of statistical test required to analyse them. The numbers we assign to a measurement, whether a questionnaire response or the reading from a technical instrument, are classified into four distinct levels referred to as scales of measurement' (Oseland, 2007, p. 66).

2.2.2 Categories

When the closed questions are made up of categories, they are generally looked at and analysed as nominal scales. This is the first, most basic, level of scales.

Nominal scales

Nominal scales are basically coding for categories, rather than a value. What can be determined from these scales is whether the items are the same or different. They do not provide a way to look at how responses relate to each other. In occupant questionnaires a nominal scale could be used for the individual occupant, e.g. male/female, or how they use the space, e.g. sit next to a window/not near a window.

2.2.3 Rating scales

Three further scales of measurement are different forms of rating scales. These are widely used in occupant questionnaires of buildings.

Ordinal scales

Ordinal scales are the second level of scales, and the first level of the rating scales. Ordinal scales allow for ranking of scores, so a response can show that one is better than another but the differences between them are not quantified. The order of the value is important but the intervals between them are not necessarily equal.

‘That is, when people choose among these ratings, the additional quantity of a characteristic that causes them to raise their rating of 1 to 2 is not necessarily the same amount that would cause them to raise their rating of 2 to 3. It is difficult to ascertain the mental scales individuals use when they express opinions as ratings’ (Aronoff & Kaplan, 1995, p. 346).

Interval scales

As well as allowing for a ranking between scores, the intervals between the scale points are equal in interval scales. The distance between the variables in this type of scale has meaning. The most common form of interval scale is for variables such as temperature using a Celsius or Fahrenheit scale. The zero point on this scale however, is arbitrary so ratios between values are not meaningful.

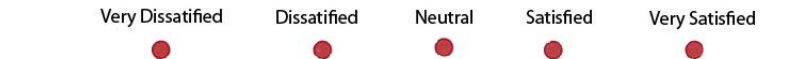


Ratio scales

This level of measurement has the same properties as an interval scale as well as having a fixed zero point. This means that differences between the scores as well as relationships and ratios between the scores are meaningful. These scales represent the highest level of measurement so have the most flexibility in terms of what statistics can be used with them, and can be determined largely though user preference.

Examples of rating scales

There are many different types of scales that can be used that fit within these three categories. Oseland (2007) in his ‘Guide to Post-Occupancy Evaluation’ identifies a selection of these scales which are in the following table.

Table 1: Summary of common scales (Oseland, 2007, p. 32)

Dichotomous Scale	Only two choices, e.g. 'agree' or 'disagree'
Thurstone Scale	Where respondents agree or disagree with a series of statements (a precursor to the Likert scale)
Labelled/Categorical Rating Scales	Multiple choice answers to a question in the form of a series of labelled tick boxes, which are usually progressive and symmetrical 
Likert Scale	A method of measuring attitudes, where respondents are allowed a degree of disagreement by being offered a rating scale ranging from 'strongly agree' to 'strongly disagree.' The range of responses on the Likert Scale was presented as a five-point labelled scale; owing to its popularity all five-point labelled scales are now commonly referred to as Likert scales 
Bipolar and Semantic Differential Scales	A bipolar scale offers respondents a numerical range, with labelling at either end of the scale only. A semantic differential scale is, typically, a seven-point bipolar scale where the end points are labelled with adjectives or phrases of opposite meaning, e.g. 'very dissatisfied' to 'very satisfied' 
Rank Order Scales	The respondent is asked to put a range of options in order of preference or evaluation of a certain attribute or characteristic. For example, a question may be 'rank the following environmental conditions in order of importance' followed by various responses, such as 'temperature', 'ventilation', 'noise' and 'light'. The respondent assigns a number to each response to represent the rank order of the item (ties may or may not be allowed).

The main difference between the 'Labelled/Categorical Rating Scales' and the 'Likert Scale' is the wording of the categories. When the 'Likert Scale' was created it was specifically in regards to the 'disagree'/ 'agree' labels. In more recent times, researchers apply the term more frequently to a broader range of five point response scales.

These scales fit into different scales of measurements so there will be different statistical tests appropriate for each. The decision of the appropriate tests to run is also not always cut and dried as some of these scales can fit into different rating scale categories and it is at the discretion of the researcher as to in what category they are assessed.

Points on a scale

There is then the added choice of how many points the scale should have to achieve the appropriate results. There are many different arguments about this and generally it comes down to personal preference, and the scope and intent of the questionnaire. Researchers have chosen to use different numbers of points on a scale over the years in questionnaires to assess the indoor environmental quality of buildings.

One of the main differences between scales is whether they are made up of an odd or even number of options. If they are even then respondents are forced to make a judgement that it is either positive or negative, where they may be more inclined to express a neutral rating.

'For many environmental features, a neutral rating is actually a good indicator and a desirable score. People generally perceive their physical work setting as a backdrop to their activities. It tends to be noticed only when it is negative or objectionable' (Aronoff & Kaplan, 1995, p. 348).

Therefore, many researchers who have carried out questionnaires of environmental quality of buildings recommend the use of an odd numbered scale to allow for this neutral option.

'The way rating scales are labelled and the number of values in the scale greatly affect both the internal and external consistency of the data' (Aronoff & Kaplan, 1995, p. 347).

The respondents need to be able differentiate between the points on the scale, while simultaneously being able to place themselves within it (Moser & Kalton, 1993). Miller (1956) identified through his own experiments as well as others that people are able to differentiate between seven items before they start to get confused. Oseland (2007) then concluded that this makes seven a sensible maximum number of points on a rating scale, and that a general rule of thumb for rating scales was to limit scales to 5 ± 2 points.

2.3 Examples of questionnaires used

There are many types of questionnaires that have been used for assessing the indoor environment. A sample of these was studied to determine what established methods are in use and the variables that are used to compile an effective questionnaire. Their main features will be summarised in the following sub-section.

There have been previous literature reviews on different methods of assessing the indoor environmental quality from the occupants' perspective (Oseland, 2007; Usable Buildings Trust, 2011; Peretti & Schiavon, 2011). Oseland (2007) compiled a list of 20 regularly used POE questionnaires from the UK. The Useable Buildings Trust Feedback Portfolio (2011) summarises methods which are available for the assessment of building performance, also restricted to the UK. Peretti and Schiavon (2011) list and describe features of ten available international questionnaire based methods.

Using the findings from these reviews, and an additional literature search, a sample of 20 previously used questionnaire based methodologies was collected. The search was limited to methodologies that were for places of work, and that included an assessment of the indoor environmental quality from the point of view of the occupants using a questionnaire. Those with readily available information and that used rating scales in their questionnaires were included in this sample; if the methodology for these projects included the same questionnaire as another, the questionnaire itself was looked at and was only included once. The questionnaires often focussed on a specific aspect of the indoor environment, such as how it affected productivity or health, so there is a slight shift in focus among the sample. These 20 questionnaires were then classified and analysed in terms of the questionnaires themselves, and how the results are summarised and/or benchmarked. Table 2 below summarises the questionnaire based methods, and the questionnaire details. Further details on the questionnaire based methods can be found in Appendix A.

Table 2: Questionnaire examples

	Survey Name	Country	Year	Building Type	Physical Measurements	Main Media Used		Number of Points					Type of Scale Used	Number of Buildings Surveyed
						Paper	Web	3	4	5	6	7		
1	BOSTI (Buffalo Organisation for Social and Technological Innovation) Survey	USA	1977	Offices	No	X				X			Likert	More than 70 office sites, and between 5,000, and 6,000 responses
2	BUS Occupant Survey	UK	Since 1985	Offices, Residential	No	X						X	Semantic Differential Scales	Over 400 organizations and individuals worldwide
3	Physical Work Environment Satisfaction Questionnaire (PWESQ)	Australia	1986	Workplaces	No	X				X			Likert	NA
4	REF (Ratings of Environmental Features)	USA	1987	Workplaces	No	X			X		X		Likert	7 administrative units and offices
5	CWRE (Checklist of Work Related Experiences)	USA	1987	Workplaces	No	X			X	X			Categorical checklist, ranging from 5=always to 1=never	7 administrative units and offices
6	AMA WorkWare	UK	1990	Offices, Education	No		X			X			Labelled/ Categorical Rating Scales	300 buildings - 65,000 occupants
7	OLS (Overall Liking Score)	England	1992	Workplaces	No		X					X	Semantic Differential, Double Likert, BSI	100 in ABS database
8	Building Assessment Survey and Evaluation (BASE)	USA	1995-1998	Offices	Yes	X			X	X			BSI, Check boxes, Categorical	100 buildings in 37 cities and 25 US states
9	ProKlima	Germany	1995-2003	Offices	Yes	X				X			Likert	14 German office buildings
10	ASHRAE RP-884	USA	1996-1998	Offices	Yes	X		X				X	Likert	160 buildings, approximately 21,000 subjects
11	CBE Survey (Centre for the Built Environment)	USA	1996	Workplaces, residential, educational	Yes		X					X	Semantic Differential Scales	600 buildings. Approximately 60,500 subjects in US, Finland, Germany, Mexico, Italy
12	SCATS (Smarts Control and Thermal Comfort)	5 European Countries	1997	Offices	Yes	X				X	X	X	ASHRAE scale and Likert scale	26 buildings in England, Sweden, Portugal, Greece and France. Approximately 4650 subjects
13	HEFC(Higher Education Funding Council) methodology	UK	1998	Universities	Yes	X	X					X	Semantic Differential Scales	NA
14	DQI (Design Quality Indicator)	UK	1999	Offices, Education	No		X				X		Categorical ranging from disagree to agree, with not applicable as an option	100 organisations
15	OPN (Office Productivity Network)	UK	1999	Workplaces	No	X	X			X			Labelled/ Categorical Rating Scales	SHC - 70 buildings - 7000 responses, ipd - 1800 responses
16	TOBUS (Tool for Selecting Office Building Upgrading Solutions)	5 European Countries	2000	Offices	No	X	X		X				Labelled/Categorical Rating Scales	15 buildings in France, The Netherlands, Switzerland, Greece, Denmark

	Survey Name	Country	Year	Building Type	Physical Measurements	Main Media Used		Number of Points					Type of Scale Used	Number of Buildings Surveyed
						Paper	Web	3	4	5	6	7		
17	COPE (Cost-effective Open Plan Environments)	Canada	2000	Offices	Yes		X					X	Likert, agree/disagree, rank order scale	9 buildings
18	HOPE (Health Optimisation Protocol for Energy-efficient Buildings)	9 European Countries	2002	Offices, Apartments	Yes	X						X	Semantic Differential Scales, Likert	164 buildings in 9 European countries (69 were offices and 95 apartment buildings)
19	WEDI (Work Environment Diagnosis Instrument)	Netherlands	2003	Workplaces	No		X			X			Likert	23 case - 7,000 respondents (71 cases - 10,000 responses)
20	RPM (Remote Performance Measurement)	Denmark	2004	Offices	Yes		X					X	Likert	Approximately 30 buildings, 1500 people. Now is in use for research purpose

The questionnaires in the analysed sample were from around the world, ranging from the Netherlands to Australia. Some of them were carried out in one-off studies, while others have been used consistently over a number of years and a range of building types. All of them have been developed over many years; the earliest looked at was from 1977, while the most recent 2004. The spread of when these were developed show that the indoor environment has been a point of interest for many years and that questionnaires have proved to be a popular tool to evaluate the building environment.

The majority of the questionnaires were developed to be used as an assessment of offices, although there were some for defence and education buildings.

Six of the projects undertook physical measurements in the space as well as the occupant questionnaire, while three additional projects had the option of physical measurements depending on the project. The structure of the questionnaires was largely related to the objectives of each project. There were several sections related to demographics, SBS symptoms, comfort and job satisfaction.

The questionnaires employed different methods of presenting the questions to the occupants. 50% of the questionnaires were predominantly paper-based, 35% were web-based and 15% were developed so the researcher could have a choice in which they used. There was also one option where the web-based questionnaire was presented to the occupants on a hand-held computer. Some of the media were determined due to the era in which they were developed and the availability of resources, e.g. paper rather than web-based. However, this was not the case in all of them. An important factor that many of the researchers found was that the response rate in paper-based questionnaires was found to be higher. An example of this is the Office Productivity Network (OPN) questionnaire where the client is offered the choice of either a paper or web-based questionnaire. The response rate of the web-based is generally around 35-50% while the paper-based is much larger at around 70-85% (Swanke Hayden Connell Architects, 2005).

The questionnaires all employed rating scales to allow for quick response and easy analysis; however the number of points on the scales differed. The analysis of the various questionnaires generally considered the most frequently used scale for that study. In **Section 2.2.1: Questionnaires as a tool** it was found that the best scales were generally 5-point \pm 2, and looking at the range of the scales used by these projects it can be seen that they have all used this rule of thumb, as the shortest scale is 3-point and the longest 7-point.

In this sample the large majority of the questionnaires were made up of scales with odd numbers of points, allowing for the occupants to make a neutral rating. There were equal numbers of scales with 5-points and 7-points, which made up 70% of the questionnaire sample. This proportion of the different numbers of questionnaires utilising different scales was different to what Oseland (2007) found in his UK sample. In his UK sample most of the questionnaires utilised 5-point scales.

Oseland UK Sample

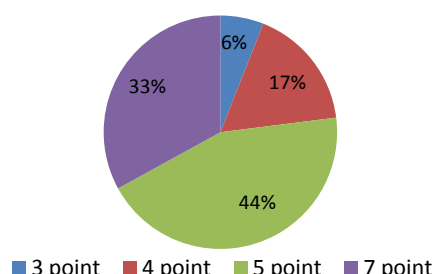


Figure 3: Types of scales in Oseland's sample

20 studied questionnaires

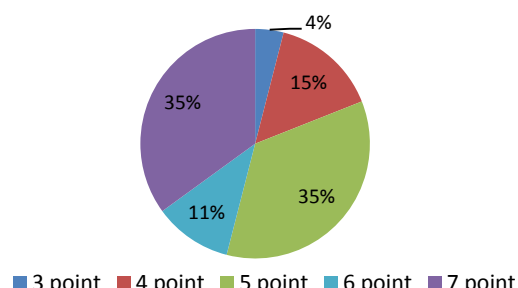


Figure 4: Types of scales in 20 questionnaires studied

As well as the differing number of scale points, the questions and rating scales were worded differently, mainly due to the researchers' personal preference during the questionnaire design stage. Using the definitions of the different rating scales in **Section 2.2.1: Questionnaires as a tool**, there was a mixture ranging from 'Likert Scales' to 'Labelled/Categorical Rating Scales' to 'Bipolar and Semantic Differential Scales'. For example the Center for the Built Environment (CBE) Occupant Indoor Environmental Quality Survey uses a 'semantic differential scale', classified by the bipolar labels placed at either end, for example:



Figure 5: Example semantic differential scale (Center for the Built Environment, 2009)

The OPN survey used 'labelled/categorical rating scales' phrased as follows:

Environmental conditions

How satisfied are you that the following environmental conditions support your work activities?

(tick one circle per environmental condition)

	Very dissatisfied	Dissatisfied	Indifferent	Satisfied	Very satisfied
Indoor temperature in winter	01 <input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 6: Example labelled/categorical rating scale (Swanke Havden Connell Architects, 2005)

Although these are essentially rating the same thing, the CBE scale allows more interpretation by the occupants due to the fact that each point is not labelled. It is assumed that each point on the scale equals an increase of equal proportion of dissatisfaction, whereas the OPN explicitly labels each rating.

Some of the questionnaires have been used more extensively than others. The most extensively used questionnaires that were looked at were the Center for the Built Environment (CBE), Building Use Studies (BUS) and the Overall Liking Score (OLS). Using the CBE as an example, this has been used in up to 600 buildings, while some of the others were used for specific one-off studies and hence do not have a large database of buildings surveyed using the same methodology. The OPN and the Alexi Marmot Associates (AMA) questionnaires have also been employed in many buildings in the UK. These, however, have been carried out by different organisations so it is difficult to determine the total number of buildings surveyed. For example, OPN has been used by both Swanke Hayden Connell Architects and Investment Property Databank (IPD) (Building Use Studies, 2010). Each company has their own database and hence have their own set of benchmarks.

2.4 Background summary

This chapter aimed to provide a background to the study; why the indoor environment is important to occupants, and ways to determine occupants' perceptions. There has been a large amount of research on the indoor environment of buildings, and the significant affect it can have on the occupants, in terms of their health, comfort, and productivity. Questionnaires have been determined to be an effective tool for assessing the indoor environmental quality of buildings; providing an assessment of how the occupants feel the building is performing.

Examining previously used questionnaire based methods allows an understanding of when, why and how subjective tools can be used for IEQ analysis (Peretti & Schiavon, 2011). There are general rules of thumb for questionnaire based methodologies, but if general good practice is followed a high response rate can be achieved.

3.0 BENCHMARKS

The objective of this chapter is;

- To explore the nature and characteristics of benchmarks in general; those in which people's perceptions play a key part, and those already applied in the building industry.

This chapter is made up of six individual sections, each with a specific focus:

- Section 3.1 outlines the background to benchmarks in general, and the differing ways they can be used and established.
- Section 3.2 is an investigation of benchmarks that are already established in the building industry, with a particular focus on energy benchmarks.
- Section 3.3 focuses on benchmarks that are constructed from results from occupant questionnaires.
- Section 3.4 is a more in depth analysis of the BUS benchmarks made from the BUS questionnaire which is the focus of this research.
- Section 3.5 discusses current building sustainability rating tools.
- Section 3.6 provides a summary discussion on benchmarks.

3.1 Nature and characteristics of benchmarks

To explore benchmarks of building performance it is first necessary to consider the nature and characteristics of benchmarks, and the interpretations currently offered both in the building industry and in more general terms.

'Benchmarking is a process of comparing the characteristics of one or more systems to a set of standard measurements, known as benchmarks. Fundamentally, the objective of benchmarking is to evaluate the performance of a system by comparing selected measurements of its status or operating characteristics to a suitable set of benchmark values. The benchmark measurements are thus indicators of performance' (Aronoff & Kaplan, 1995, p. 377).

Benchmarks and benchmarking are used through many industries in different ways. There are many differing ways of defining a benchmark. One of the most common practices of benchmarking is in the management field. Although not the same field as the focus of this study, there are many similarities, and the ideas behind the benchmarking are the same. In management, benchmarking is used to compare how different companies are managed and whether any of the techniques they use are able to be usefully transferred into other businesses. It can be used as an investigative tool to study differences in performance levels.

The most generic definition comes from the Oxford Dictionary, which defines a benchmark as 'a standard or point of reference against which things may be compared' (Oxford University Press, 2010).

Roaf (2004) looks at a benchmark in the building industry and compares an indicator to a yardstick.

'In the context of the built environment, for any attribute of a building (e.g. its energy use, noisiness, comfortableness) there is a corresponding yardstick to measure how the building performs in relation to that attribute or issue' (Roaf, 2004, p. 35).

Dziegielewski (2000) relates benchmarking in the management field to benchmarking water use in buildings.

'Benchmarking is a method used by businesses to measure their performance relative to the performance of other businesses. Benchmarking can also be used to assess the efficiency of water use in a business establishment' (Dziegielewski, 2000, p. 116).

Benchmarking allows comparisons between buildings and can be used for many different aspects in a range of industries, including the building industry.

Although benchmarking and benchmarks have a range of specific meanings, mainly in relation to the industry or topic they are relating to, there is the underlying commonality that a benchmark is a feature of improvement projects that allows a comparison.

Benchmarks are important as they set a standard which something can be compared to, or aimed for. Allowing for this comparison they allow a single building, organisation, task etc. to be looked at in relation to a large dataset. They have become more used in the building industry in recent years due to the increasing trend for sustainability and energy efficiency. The most obvious examples of this are the building rating tools that many countries have developed, such as Green Star in NZ, LEED in the USA, and BREEAM in the UK. This will be considered in further detail later in **Section 3.5: Current building rating tools**.

Roaf (2004) offers some guidance on how to choose the right indicators and benchmarks. Her guidance is directed to those who are selecting appropriate benchmarks for their particular enquiry but is also useful to those involved in developing the benchmarks themselves:

- **Understand the source:** this involves knowing where the benchmarks come from and assessing whether they have clear, unbiased agendas, and whether they are based on predicted or actual performance.
- **Aggregated indicators:** this compares the advantages and disadvantages of a single index to an overall profile.
- **Compare like with like:** there needs to be a clear definition of an indicator, so poor performance is not hidden.
- **Choose the right scale:** this considers the effect of the different categories of a scale and the difference it can make to the overall assessment.
- **Think about the user as well as the source:** this considers how the assessment is presented to the client.
- **Learn from experience:** use lessons to improve building performance and reduce related impacts over time (Roaf, 2004, pp. 41-48).

As a benchmark can be looked at as a point on a line, the issue becomes what is the best way to determine the length of the line and where the specific benchmark value fits along the line. Throughout the building industry there are many different theories of what the 'best' is. Benchmarks can be presented in different forms. While some people prefer figures against which comparative numbers can be evaluated, others prefer bands of performance into which their results can fit. Each of these offers different advantages and disadvantages and are mainly determined by personal preference and the type of data that has been collected. All of these techniques use the same basic statistical methods that are adapted to suit the particular needs of the project, but are applied in different ways in rating tools or comparison methods. This will be looked at further when considering benchmarking examples.

3.1.1 General statistics

When a benchmark is determined from large amounts of data, the particular statistic used is specific to the data and use of each project. Generally, the benchmarks are determined through the use of a measure of central tendency, a measure of variability, and a measure of relative standing.

Data distribution

When a sample is collected, the data received is variable and can be distributed in many ways. The most commonly observed distribution is the 'normal distribution.'

'The normal distribution is a mathematical function widely used in statistics to calculate the probability of events. It produces the familiar bell-shaped curve known to everyone exposed to introductory statistics' (Aronoff & Kaplan, 1995, p. 353).

This is a symmetrical bell shaped curve, an example of which is illustrated below in Figure 7. Most of the area under the curve falls in the centre. The tails of the distribution approach the x-axis but never touch, with minimal area under them.

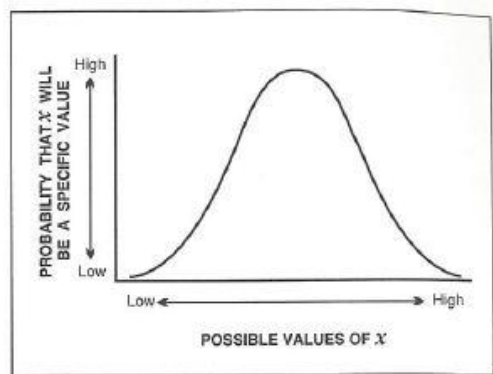


Figure 7: Normal distribution example (Aronoff & Kaplan, 1995, p. 353)

Many statistical analyses rely on the assumption that the distribution of the data reasonably approximates a normal distribution.

Central tendency

There are different ways of determining the measure of central tendency of a distribution. These are most commonly the mean, the median and the mode.

The **mean** is the most common and is calculated by dividing the sum of the scores by the number of scores. This often gives a good interpretation of the central tendency of the scores, but can be misleading in the case of extreme scores or outliers. If there are extreme scores, these will bring the mean up or down and may result in the mean not being representative of the distribution of scores (Salkind, 2005). The data needs to be quantitative for this to be calculated. In terms of rating scales, this can be used with interval and ratio scales.

The **median** is defined as the midpoint of a set of scores. That is, half the scores fall above and half the scores fall below this point. The most advantageous aspect of using this average is that it is insensitive to extreme scores. In cases where there are extreme scores it better represents the centremost value than any other measure of central tendency (Salkind, 2005). It is, similar to the mean, applicable to quantitative data only. The median can be used with data from ordinal, interval and ratio scales.

The **mode** is most commonly used when the data is categorical. The mode is then the category that appears most often. There can be more than one mode in a data set. It is not influenced by extreme measurements. It is applicable to both quantitative and qualitative data. The mode can be used with all types of scales, and is most often used with the nominal scales.

These measures all have a relationship to each other, which changes depending on the skewness of the data. If the sample is normally distributed, the mean, median and mode will all coincide in the same place, the centre of the distribution (Ott & Longnecker, 2010).

These measures present the results in the most manageable form - a single number. This, however, limits the amount of information that is being portrayed, and means that the original data is not able to be reconstructed. This can in some ways be lessened by also presenting the summaries of variability which represent the distribution of the scores.

Variability

Three measures of variability that are commonly used are the range, the standard deviation and the variance. Measures of variability are used in conjunction with a measure of central tendency, some of these being specific to the type used.

The most general measure of variability is the **range**. This gives an idea of how far apart scores are from one another. It is calculated by subtracting the lowest score from the highest. It is used mainly as a general indicator of variability and does not portray how individual scores differ from one another (Salkind, 2005). The range is more informative as a measure of variation for small sets of data, than large sets.

The **Standard Deviation** (SD) is the most often used measure of variability. It is a measure of the degree in which the data is spread around the mean, and represents the average amount of variability in a set of scores. A higher standard deviation means the data is spread out, while a smaller standard deviation indicates the scores are closer together around the mean. The standard deviation is also sensitive to extreme scores.

It is calculated so that for a normal distribution about 68% of the scores will lie between -1 and +1 standard deviation from the mean, 95% will lie within -2 and +2 standard deviations and 99.9% will lie between -3 and +3 standard deviations.

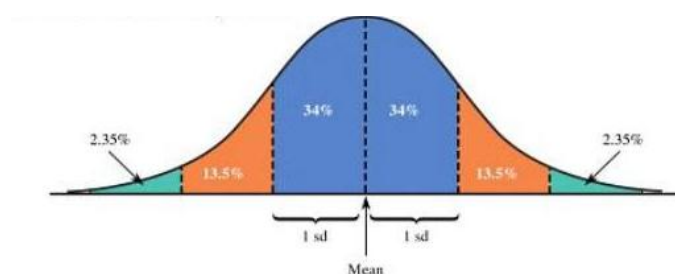


Figure 8: Standard deviation example (Peck, Olsen, & Devore, 2009, p. 180)

The **variance** is simply the standard deviation squared. The benefit of this is that it is an absolute number; there are no positive or negative values.

Measures of relative standing

Once you obtain a rating, there are several statistics that can be used in order to know how it compares to the rest of the sample. The statistics are measures of the position of a particular value in a data set relative to all values in the set.

One of these measures is the z-score. The z-score measures how many standard deviations the value is from the mean. It is positive or negative according to whether the value lies above or below the mean. The z-score is particularly useful when the distribution of observations is approximately normal. In this case, a z-score outside ± 2 occurs approximately 5% of the time, and outside ± 3 only about 0.3%. (Peck, Olsen, & Devore, 2009). A z-score is calculated by subtracting the mean of the distribution from the score, and then dividing by the standard deviation, as shown in Equation 1. Z-scores can be calculated for interval and ratio scale data.

Equation 1: Standard score

$$z = \frac{(X - \bar{X})}{s}$$

Another measure that can be used is percentiles. This is when a value can be located by giving the percentage of data that falls at or below that value, when ordered according to magnitude. For example, the 60th percentile for the variable x is a point on the horizontal axis of the distribution that is greater than 60% and less than the others; so 60% of the measurements are less than the 60th percentile and 40% are greater. The median is also a percentile; as it is the half-way point in a distribution it represents the 50th percentile. There are several other specific percentiles that are commonly used. Percentiles are able to be used with ordinal, interval and ratio scale data.

When you order the data in ascending order and split it into even groups, these groups are referred to as quantiles. These can be split into any number of bands, but a common quantile that is used is the quartile. This is where the data is split into four even groups, with each band being worth 25% of the population, with the 2nd quartile mark being equivalent to the median. The upper and lower quartiles are equivalent to the 25th and 75th percentiles.

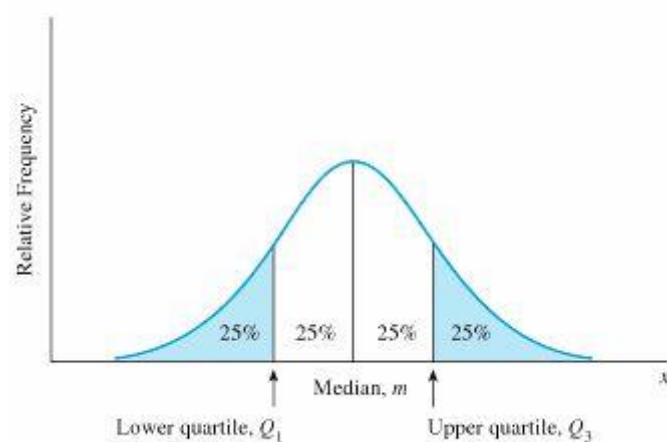


Figure 9: Quartile example (Mendenhall, Beaver, & Beaver, 2009, p. 77)

3.2 Benchmarks already applied to the building industry

Benchmarks are fundamentally the same, independent of what they are assessing. They are often formatted in different ways, to suit the various data being analysed. However they are all used as an indicator of how something is performing, allowing a comparison. They are used throughout the building industry as well as many other industries. Buildings are being studied more frequently due to the increased importance placed on the cost and performance of a building. Building benchmarks are being used as part of these studies.

One of the most publicised benchmarks that is used in the building industry is energy. For this reason energy benchmarks are used in this study as a general example of benchmarks and their application in the building industry.

3.2.1 Energy benchmarks

In recent years energy benchmarking has become more important as there is an increased focus on energy efficiency. 'Benchmarking energy-efficiency is an important tool to promote the efficient use of energy in commercial buildings' (Chung, Hui, & Miu Lam, 2006, p. 1).

There are two fundamental purposes of benchmarks of energy use. The first is to identify how a building's energy-use rates, compared with other buildings of its type. In this case, 'empirical benchmarks derived from energy statistics for the stock are applicable' (Cohen, Bordass, & Field, 2006, p. 2). The second purpose is to determine whether a building is performing to its potential, and how much it might be improved. In this case, 'a realistic model of the building is theoretically more applicable' (Cohen, Bordass, & Field, 2006, p. 2). As this project is focussed on identifying how a building rates it is the first of these purposes and hence the empirical benchmark that will be looked into further.

Perez-Lombard et al (2009) discuss the difficulties of determining an energy benchmark. There are four stages in the benchmarking process.

'First, it is necessary to hold or develop a database with information on the energy performance of a significant number of buildings. This information should be categorised, at least, by building type and size. Second is gathering the relevant information for the evaluation of the EPI [Energy Performance Index] for the actual building. Third, a comparative analysis of the building energy performance against the samples held in the database gives a quantification of the quality of building in terms of energy use. Finally, energy efficiency measures that are feasible from both technical and economical perspectives should be recommended' (Perez-Lombard, Ortiz, Gonzalez, & Maestre, 2009, p. 274).

Database

The four step process described above is applicable to most benchmarking practices, including user perception benchmarks which are the focus for this study; however as it is focussed on the benchmark itself, the fourth stage of improvement will not be investigated further in this study. The improvement of the performance of commercial and institutional buildings is still the overall aim, and establishing benchmarks is a necessary precursor to further work.

The most difficult of the steps stated above is the first. Gathering enough information to fill a database with information representative of the building stock is complex and expensive. Because of these difficulties, not many countries have a reliable set of measured energy data for much more than a small sample of buildings. An excellent example is the US Energy Information Administration (EIA) database (Perez-Lombard, Ortiz, Gonzalez, & Maestre, 2009). The process in the example of energy use benchmarking is very similar to what can be used for benchmarking user perceptions, so obtaining a representative sample in terms of user perceptions will also face difficulties.

In energy benchmarking, a summary of each building's energy use needs to be calculated to make up the database. This is commonly referred to as an Energy Performance Index (EPI) (or Energy Use Index (EUI)), and this is generally normalised according to floor area so comparisons between buildings of differing sizes can be made.

'Some authors propose multiple indices to consider simultaneously energy use, environmental impact and indoor air quality, though energy use per unit of area and year is almost the standard EPI for buildings' (Perez-Lombard, Ortiz, Gonzalez, & Maestre, 2009, p. 277).

Once it is determined how the building's energy use will be summarised, the type of buildings that will be included also needs to be considered. Cohen et al (2006) suggest that there can be difficulties in 'getting the contextual data right, in terms of classification of the building, its systems and its use.' These issues will be carried through to user perceptions, and were briefly mentioned in **Section 2.1: How the building environment affects the health, comfort, and productivity of building occupants**. Perez-Lombard et al (2009) also suggest that the buildings must be comparable and that the 'minimum degree of similarity' is two: same climate and building type. This indicates for example that the different buildings should be looked at individually before they are included in the database.

The final issue is the quality of the data. Cohen et al (2006) state that in their experience many records at either end of the statistical distribution include faulty data, resulting in outliers. This is not such a big problem when the sample is large. They are of the opinion that this problem means that the median is a more robust indicator than quartiles, deciles or averages, as the median is not influenced by outliers while the mean is (Cohen, Bordass, & Field, 2006).

Comparisons

In terms of the third step, the Empirical Benchmark can be broken down further, by the way the comparisons between buildings are made. The simplest, and most conventional, is obtained from bulk statistical data, and is often a single statistic that allows for a direct comparison. The second is also obtained from statistical data but is parameterised. This allows for buildings to be placed within different levels of set criteria. This is used relatively frequently particularly within building rating tools (Cohen, Bordass, & Field, 2006).

There are many examples for both of these methods. One example for the direct comparison of a single statistic is the Energy Star program. A single number is often used for a general target benchmark, or a minimal set level. The Energy Star program has been used to benchmark energy use in buildings. The primary data sources for this were data collected by the US Bureau of the Census. This program uses a percentile rating system from 1 to 100, for specific commercial building types. To achieve an Energy Star rating the building has to achieve a minimum of 75 points, which is in the top quartile of energy efficiency. So in this case quartiles were used to benchmark their energy use (Perez-Lombard, Ortiz, Gonzalez, & Maestre, 2009).

In the project carried out by Hernandez et al (2008) for developing benchmarks for non-domestic buildings using Irish primary schools as an example, they demonstrate, along with previous projects, that the distribution of specific energy consumption is likely to be positively skewed rather than normally distributed. They argue that for benchmarks in this situation the median is more appropriate as a measure of typical building stock performance rather than the mean. However, the mean and standard deviation are still used to determine outliers. It will be interesting to see how this compares with the data of user perception and whether the same statistical concepts can be applied.

In terms of examples of parameterised benchmarks, there are many to select from. For example, in 'Energy Consumption Guide 19: Energy Use in Offices' (ECON 19), which is a guide developed by Building Research Establishment (BRE), there are 'Good Practice' and 'Typical' standards for each energy use in building. The typical value is the equivalent of the median of the building stock, while the good practice rating is the lower quartile (Action Energy, 2000).

The main determinant for parameterised benchmarks is where the splits should be. Cohen et al (2006) discuss the benchmarks for operational ratings. For the simple benchmark, the data for a particular building type is usually divided into bands containing equal number of buildings, e.g. many UK publications for benchmarking buildings have used quartiles. This is then looked at in terms of performance. Those in the best quartile are often termed 'good' practice while those in the worst as 'poor' (Cohen, Bordass, & Field, 2006).

Perez-Lombard et al (2009) consider building energy labelling. Building Energy Certificates have a list of minimum requirements. One of these is to include a label base from A-G to achieve a suitable rating of a building. The definition of the scale needs to make reference to the building energy regulation, the existing building stock and the zero-energy building. If there are enough comparable buildings, statistical analysis of the Energy Performance Indicator (EPI) by using cumulative frequency distribution will allow the percentile to be an indicator of the energy position. This then becomes a case of applying percentile intervals (bands) to energy classes (Perez-Lombard, Ortiz, Gonzalez, & Maestre, 2009).

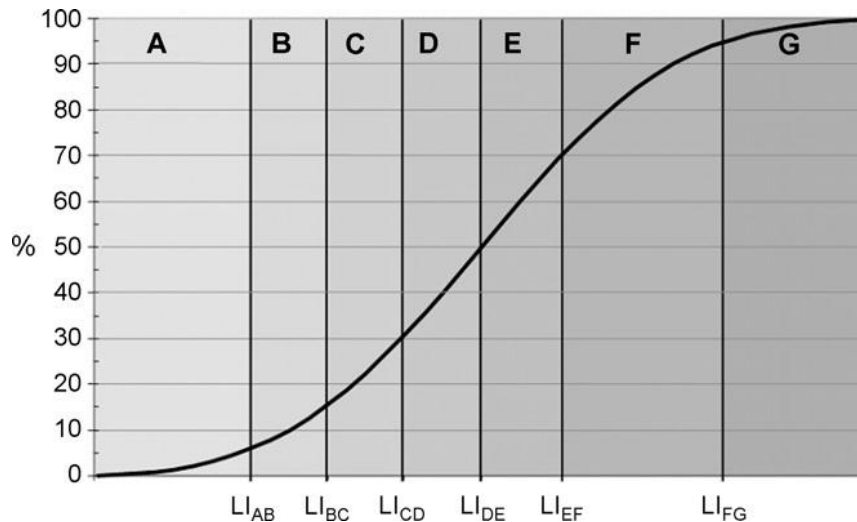


Figure 10: Labelling scale and cumulative frequency curve (Perez-Lombard et al. 2009, p.276)

Santamouris et al (2007) argue that for a dataset of buildings, which have a variety of characteristics, the data will often not follow a normal distribution. A normal distribution is needed for the classification of the quartiles of the cumulative distribution that is currently used. They put forward the idea that in most cases the existing energy data is combined around various clusters that may not be represented by a normal distribution. In this case it is appropriate to use cluster analysis;

‘a mathematical procedure to identify natural groupings of objects, in such a way that the characteristics of objects belonging to the same cluster are very similar while the characteristics of objects in different clusters are quite distinct, producing thus a concise representation of the dataset behaviour’ (Santamouris, et al., 2007).

From the analysis of the energy benchmarks there are several main points that should be carried through and kept in mind when assessing the user perception data in the benchmark development process.

- It is important to take into account the classifications of the buildings that are being assessed.
- It is also important to look at the distribution of the data in the development of benchmarks; whether the data is normally distributed or not will affect the selection of an appropriate statistic to use.
- Benchmarks take many formats and they all relate back to the specific use of the benchmark, and the data gathered.

3.3 Benchmarks of people's perceptions

Numerous studies have looked into benchmarking people's perceptions over a range of subjects. This section will include a brief summary of how other international studies have benchmarked the results from user questionnaires of buildings. This section follows on from **Section 2.3: Examples of questionnaires used**. The same sample of 20 questionnaire based methodologies is studied but with the focus on the questionnaire data used to summarise the performance, how the results are presented, and whether the results are benchmarked against a larger dataset. This is presented in Table 3 below.

Table 3: Questionnaire based methods' benchmarks

	Survey Name	Number of Buildings Surveyed	Comparisons			Summary Statistic	Database Statistic
			No Comparison	Comparison with same study	Compared to existing database		
1	BOSTI (Buffalo Organisation for Social and Technological Innovation) Survey	More than 70 office sites, and between 5,000, and 6,000 responses		X		Mean and percentage scoring different questions	Mean and differences between questions
2	BUS Occupant Survey	Over 400 organizations and individuals worldwide			X	Mean, 3 indices, rating scores	Mean - 95% confidence intervals, percentiles
3	Physical Work Environment Satisfaction Questionnaire (PWESQ)	NA		X		Mean	Z-scores
4	REF (Ratings of Environmental Features)	7 administrative units and offices		X		NA	NA
5	CWRE (Checklist of Work Related Experiences)	7 administrative units and offices		X		NA	NA
6	AMA WorkWare	300 buildings - 65,000 occupants			X	% or overall satisfaction	Mean, range, quartiles
7	OLS (Overall Liking Score)	100 in ABS database			X	Weighted individual questions against importance - mean	Mean
8	Building Assessment Survey and Evaluation (BASE)	100 buildings in 37 cities and 25 US states	X			NA	NA
9	ProKlima	14 German office buildings		X		Individual question responses	NA
10	ASHRAE RP-884	160 buildings, approximately 21,000 subjects	X			NA	NA
11	CBE Survey (Centre for the Built Environment)	600 buildings. Approximately 60,500 subjects in US, Finland, Germany, Mexico, Italy			X	Mean, percentage satisfied	Mean
12	SCATS (Smarts Control and Thermal Comfort)	26 buildings in England, Sweden, Portugal, Greece and France. Approximately 4650 subjects	X			Coded questions according to score, mean, 95% intervals	NA
13	HEFC(Higher Education Funding Council) methodology	NA	X			NA	NA
14	DQI (Design Quality Indicator)	100 organisations	X			Individual scores and the occupants weightings to them - mean	NA
15	OPN (Office Productivity Network)	SHC - 70 buildings - 7000 responses, ipd - 1800 responses			X	% of people satisfied	Quartiles
16	TOBUS (Tool for Selecting Office Building Upgrading Solutions)	15 buildings in France, The Netherlands, Switzerland, Greece, Denmark	X			Indices, mean, max & min	NA
17	COPE (Cost-effective Open Plan Environments)	9 buildings	X			Mean, SD, kurtosis, skewness	NA
18	HOPE (Health Optimisation Protocol for Energy-efficient Buildings)	164 buildings in 9 European countries (69 were offices and 95 apartment buildings)		X		Mean - comfort, BSI - Health	Mean
19	WEDI (Work Environment Diagnosis Instrument)	23 case - 7,000 respondents (71 cases - 10,000 responses)			X	% satisfied	overall average % of satisfied/dissatisfied
20	RPM (Remote Performance Measurement)	Approximately 30 buildings, 1500 people. Now is in use for research purpose	X			Mean, range, quartiles	NA

3.3.1 Benchmarks from examples of questionnaires used

People's perceptions can be a difficult topic to benchmark due to the wide variety of forms the results can take. The main control on the data received is through careful design of the collection tool used, in this case the questionnaire. If the data is able to be transformed into a numeric form the benchmarks can be formed in a similar way for those of the more quantitative nature, such as energy benchmarks that were discussed previously. However, what is being benchmarked always needs to be kept in mind so that what is produced is appropriate.

The indoor environment is made up of many different aspects, each with their own effect on the occupant. Some of these were previously mentioned in **Section 2.1: How the building environment affects the health, comfort, and productivity of building occupants**. These different aspects give benchmarking of the indoor environment unique issues, which were not present in energy benchmarking, discussed above. The problem of 'detail versus manageability' comes to the forefront.

'In presenting univariate and other data, you'll be constrained by two goals. On the one hand, you should attempt to provide your reader with the fullest degree of detail regarding those data. On the other hand, the data should be presented in a manageable form. As these two goals often directly counter each other, you'll find yourself continually seeking the best compromise between them. One useful solution is to report on a given set of data in more than one form' (Babbie, 2001, p. 433).

Baird et al (1996) also consider this issue in their discussion on strategies and operational concepts of building evaluations.

'Overall performance as a single score versus overall performance as a profile

This is a debate over the aggregation or separation of scores in evaluation systems that utilise scoring methods. It is appealing intellectually to have one number to symbolise the total value of a building, but the argument against this is that the figure hides more than it reveals about the qualities (plural) of the building. The alternative to a single score is a group of scores for different and interesting aspects of the building. This approach can result in a kind of "personality profile" for the building as a whole' (Baird, Gray, Isaacs, Kernohan, & McIndoe, 1996, p. 13).

The results need to be presented to building owners in a clear, easy to understand way to enable them to quickly comprehend how their building is performing for its occupants.

Not all of the questionnaires studied formed benchmarks as their focus was on the standardisation of the questionnaire, or the comparison between responses and physical measurements. However, a significant number, 60% of them, did.

The number of buildings in the datasets that each study used varied greatly. Some of the one-off studies compared an individual building to an average of the other buildings that were part of the same project and were surveyed simultaneously, while some of the more extensively used questionnaires were used by different consultants and therefore each had their own separate datasets, collected over time.

A database used to create benchmarks requires the same questionnaire tool to be used, as they need to be comparable.

'To create a database for benchmarking, however, requires the same survey methodology. It is extremely difficult to compare the results from different questionnaires and databases. Even a slight change in the wording of a question can change the meaning of that question, making it difficult to compare it with other similar ones. Furthermore, it is extremely difficult to compare the ratings of building performance made on different types of scale' (Oseland, 2007, p. 38).

From this sample of questionnaire examples, the statistics that are used, the way these statistics are displayed, and the sections of the questionnaires that are used to create the benchmarks (comparison points) were studied. The way the benchmarks are displayed to the client e.g. as a summary single digit figure, or a more detailed result, or a combination of both, were determined.

'Presenting POE results in a meaningful and useful way is a matter of presenting high level information that can be quickly and easily assimilated while appearing sufficiently detailed and credible' (Oseland, 2007, p. 36).

There are many different ways that the results can be displayed to the building owners. The questionnaire sample studied displays the results in a range of ways, and Oseland (2007) summarises many of the methods.

Questionnaire data used

Although not all of the projects compared to a larger dataset, almost all of them summarised the overall results from the buildings in some way or another.

There is a range of methods for displaying benchmarks, some of which have already been looked at, and these start with the component of the questionnaire that is used. In this sample of questionnaires there are also a wide range of methods used.

- A comparison of each question to a larger dataset – Health Optimisation Protocol for Energy-efficient Buildings (HOPE) compares the individual question averages to the mean of the UK and non-UK sample.
- An overall score for the building – OLS looks at the results from each question and its corresponding importance rating and produces an overall fingerprint score for the building.
- A visual representation of the different aspects and how they relate – OPN and Design Quality Indicator (DQI) create radial diagrams, with each point representing a different question, comparing to the quartiles of a large database.

There are also several cases where the summary questions used are in the questionnaire itself. These questions try and encapsulate the overall performance of the building.

- Cost-effective Open Plan Environments (COPE) does not benchmark with a larger dataset but does provide a summary of the results. There are two summary questions in the questionnaire, 'how satisfied are you with the overall environment' and 'how does it affect your productivity'.
- CBE questionnaire includes two summary questions looking at the overall satisfaction of the occupant with the building overall, and the satisfaction of the occupant with their personal workspace.
- AMA has something similar where the occupants are asked how they rate the building overall.

These overall questions can be looked at as a summary of the conditions for a particular aspect; however, the occupants are more likely to rate the summary variables higher than a specific issue if they generally like the building. From a study of the BUS questionnaire data comparing green buildings with conventional buildings, Adrian Leaman states:

'These results look encouraging. However, there is an important caveat. These are ratings based on 'summary' responses, e.g. they include any allowances that the occupants may make for context. When one starts to look at responses in more detail, the picture is somewhat less optimistic' (Leaman & Bordass, 2007, p. 669).

Statistic used

'POE questionnaire data is usually based on an average rating or a frequency score, and an accompanying indicator of the range of responses, e.g. confidence limits or quartile ranges' (Oseland, 2007, p. 38).

All of the questionnaire based methods studied used the mean as the measure of central tendency, with differing measures of variability. This indicates that the researchers developing the benchmarks see the rating scale results as interval data, allowing more flexibility in the statistical choices.

The measures of variability used in this sample range from quartiles, to confidence limits, to standard deviations.

In terms of the individual statistics displayed there is also a variation. The COPE study displays the results with the standard deviation, kurtosis and skewness so an overall understanding of how the scores are distributed can be achieved. Some other methods do not provide as much information, for example Tool for selecting Office Building Upgrading Solution (TOBUS), which displays the mean score and gives an indication of the range by stating the maximum and minimum score in the distribution.

Quartiles are used to give an idea of how the data from an individual building compared to a larger dataset. One example of its use is with the OPN methodology, and the quartiles are displayed visually on the radial diagram (see Figure 11) alongside the score of an individual building, to see where that particular building fits into a larger dataset. AMA WorkWare also uses quartiles as a point of comparison, stating which quartile of a larger dataset an individual result fits into.

The Smarts Control and Thermal Comfort (SCATs) project displays their results with the corresponding 95% confidence levels, so the reader could be 95% sure that the population value lies between that range.

Other questionnaire examples in this sample focus on the percentage of occupants satisfied/dissatisfied. They do this in different ways. AMA uses the score of excellent/good on their 5-point scale as an indication of satisfaction, while TOBUS looks at the other end of the scale calculating the percentage of complaints. CBE also uses a summary method to present the percentage satisfied with different variables. In this method they, on some occasions, give the average of questions related to the same aspect to make an overall rating. For example, for 'Acoustic Quality' the average of two questions 'How satisfied are you with the noise level in your workspace?' and 'How satisfied are you with your sound privacy?' is used, and the average percentage satisfied is displayed.

Indices

'Questionnaire results are usually presented as averages or frequencies but the data may also be combined to provide overall scores or indices and presented on a single chart to provide a 'fingerprint' of the building. Converting the questionnaire responses to percentage satisfied may be easier to interpret than average ratings' (Oseland, 2007, p. 7).

TOBUS is interesting as it has put forward a comfort index to display a summary of the results. The comfort section is broken down into four main issues - thermal, acoustic, lighting and air - and a separate comfort index can be determined for each of these. The occupants are asked on a 4-point scale whether they feel annoyed or experience symptoms 'often', 'regularly', 'sometimes' or 'never'. The software then takes a score of 'often' or 'regularly' as a complaint and it is these responses that are used to calculate the comfort index.

Equation 2: TOBUS comfort index (Bluyssen & Cox, 2002)

$$comfort\ index_{issue} = \sum_{aspect=1}^n \frac{percentage\ of\ complaints_{aspect}}{total\ \# \ of\ aspects\ per\ issue}$$

As many of the questionnaires in the sample were developed for diagnosis of buildings affected by Sick Building Syndrome (SBS), there are large sections dealing with the effect of the buildings on the occupants' health. These were often summarised to form a Building Symptom Index (BSI).

This is generally the mean number of symptoms reported by the occupants. This is first looked at on an individual scale creating the personal symptom index (PSI), and then looked at for the whole building level.

Humphreys (2005) looks at the practicality of combined indices of the indoor environment, based on data from the SCATs project. The result from this study is that 'ranking several buildings in order of merit, by using a combination of the several aspects evaluated by the occupants, seems unlikely to rank them in the order that would have been obtained by asking them directly about the buildings' overall merit' (Humphreys, 2005, p. 325).

Presenting results

Many of the methods presented the results for an individual building in graphical form. This is also carried through to the Building Use Studies benchmarking.

The OPN questionnaire presents the results as a radial diagram, to 'produce a footprint of satisfaction with the building,' as does the DQI.

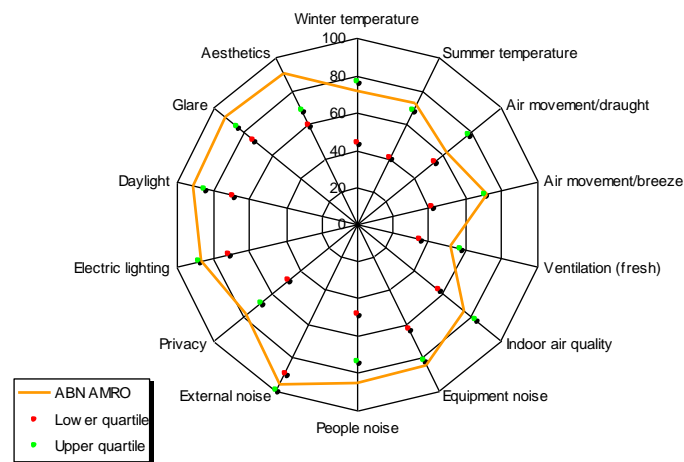


Figure 11: Example of radial diagram (Swanke Hayden Connell Architects, 2005)

The Overall Liking Score (OLS) presented the results as a 'fingerprint' of performance. As well as providing the overall fingerprint, the OLS gives the building an overall percentile score, which is an average of all the scores and this can be compared to a database average. This allows a visual representation so the client will be able to determine the high performing and low performing aspects of the building. It takes into account the rating scores and the relative importance of each variable.

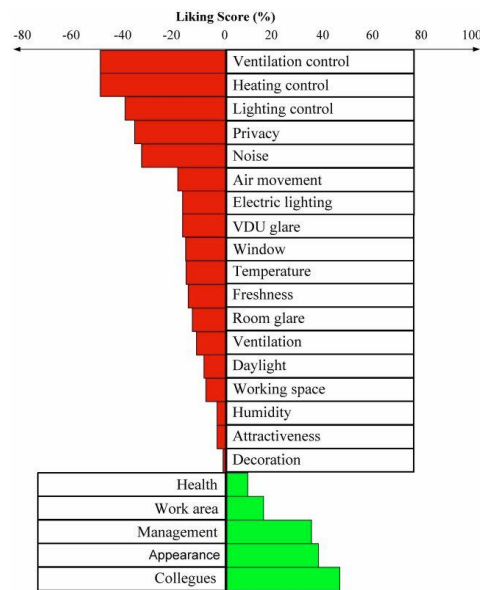


Figure 12: OLS fingerprint of performance (ABS Consulting, 2008, p. 4)

Another display option that is used is ranking the buildings in the database and reporting where the individual building fits into this, either by highlighting the individual building or splitting into quartiles/quintiles. This was shown as an example in Oseland (2007), as well as being used by AMA. The example below shows the results from a floor of a Department of Health building in London that was surveyed before and after a renovation.

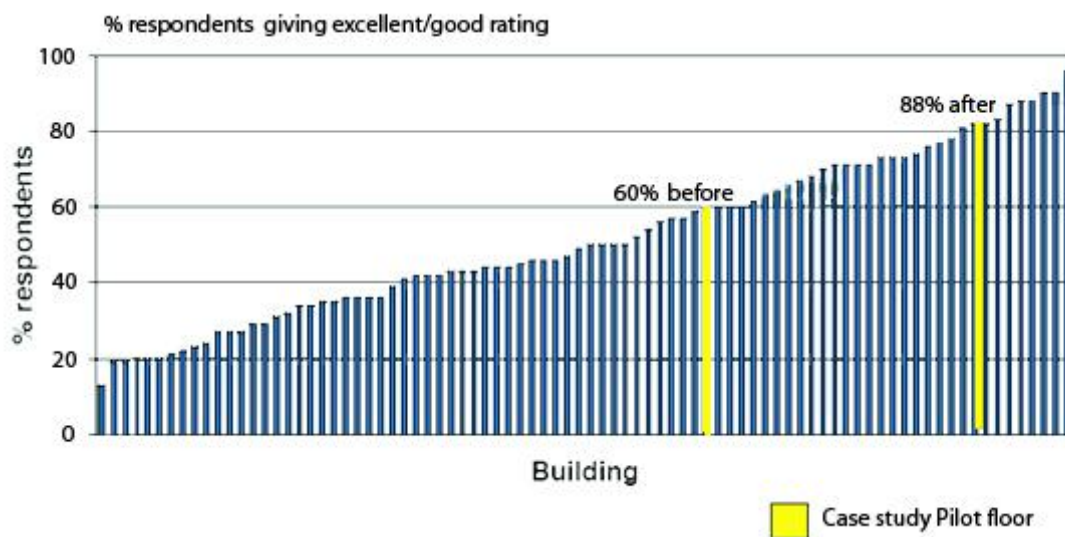


Figure 13: Ranking of individual buildings (Alexi Marmot Associates, 2004)

Oseland (2007) also includes a range of examples of ways to present the results. One of these is a simple bar chart indicating the mean range for a variety of different attributes. This shows the performance of the building in relation to different building elements, so the weak points are able to be identified.

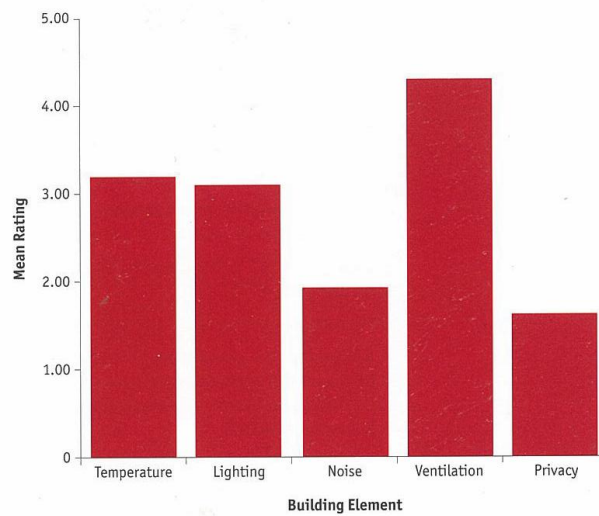


Figure 14: Example of bar chart (Oseland, 2007)

3.4 Assessment of BUS benchmarks

At present one method of benchmarking being used to rate user perceptions of building performance is that developed by Adrian Leaman of Building Use Studies (BUS). The questionnaire was developed in 1985. The benchmarks have developed and changed over the years, and there are several ways that the building performance is portrayed. These are specific to his questionnaire data and as it is this questionnaire being used in this project, the way the data is used to form benchmarks is an important area to analyse.

The benchmarking is mainly completed in two ways; looking at each question individually and comparing it to the database for an individual question, or combining a number of questions relevant to a particular topic and comparing that to the database.

Oseland (2007) identifies several methods that can be used in benchmarking and reporting of results, some of which were mentioned previously in **Section 3.3: Benchmarks of people's perceptions**. These include the formation of indices, ranking of buildings, overall classification and a band system. Building Use Studies, in their benchmarking format, use all of these in the reports that they provide the occupants. They carry out a full range of analysis and reporting techniques to provide the occupant with a very detailed analysis of how their building is performing in comparison to a dataset.

At the individual question level, the mean score of a building for each question is compared to the mean of the scores of the last 50 buildings surveyed using this POE questionnaire. These buildings are not a random sample but are determined by who was willing to have a POE carried out on their building. As these are mainly buildings where the building owner asked to be evaluated, they are often the higher performing buildings, rather than a representative sample over the whole stock of commercial buildings.

'There is some evidence to show that this set of buildings may be better than the norm, but it is difficult to test this properly without a fully randomized design' (Leaman & Bordass, 2007, p. 673).

The benchmarks used are dynamic, as they are being continually updated, so the included buildings change depending on when the surveys were conducted. This reflects what was stated in **Section 2.2: Methods of assessing the indoor environmental quality**, in which a previous study concluded that the relationship between buildings and occupants is dynamic. The BUS database is updated annually, so the buildings that are surveyed in the same year have consistent benchmark scores. The number of the buildings remains the same throughout, as the most recent 50 are taken as the benchmark. The range of movement has been shown to be consistently small over the buildings. According to Baird who has carried out many of these POEs:

‘As such, each benchmark score may be expected to change over time as newly surveyed buildings are added and older ones withdrawn. Nevertheless none of them was observed to have changed dramatically over the seven years or so during which the author has used this survey instrument’ (Baird, 2009, p. 1080).

BUS currently has separate benchmarks for the UK, Australia, international sustainable buildings and currently a tentative NZ benchmark, consisting of 24 buildings surveyed using this methodology (Baird, 2009).

Not all of the NZ buildings that have been surveyed were compared to the NZ benchmark as it has only been used in the past couple of years, since 2007, when the database was judged large enough to allow for this comparison. Earlier, the benchmark composed of the last 50 buildings surveyed in the UK was used. The variation between these will be further investigated in **Section 5.2: Comparison of benchmarks and databases**.

The buildings are compared to the database mean and are rated as being ‘Better’, ‘Similar’ or ‘Worse.’ They are considered ‘Better’ or ‘Worse’ when they are outside the 95% confidence levels of the benchmark and ‘Similar’ when within the range. In more recent surveys Leaman presents the scale mid-point as well, so this is taken into account when determining differences between the building scores and the benchmark.

Confidence intervals are used to emphasise that the data is based on sample statistics and therefore subject to variation. Building Use Studies use a confidence level of 95%, ‘given these statistical limits, we can be confident that 19 times out of 20 (95% of cases) the true value falls within the bands’ (Leaman, 1997, p. 38).

Different ways of presenting this information have been developed over the years. The most recent display format is a bar graph of the distribution of scores for each question, a slider showing where the mean scores fit compared to the benchmark and a graph showing where the building fits in regards to the buildings making up the benchmark.

The bar graph showing the distribution of scores allows the percentage of people satisfied/dissatisfied to be calculated, and allows for an assessment as to the distribution and whether there are outliers that may affect the calculated mean.

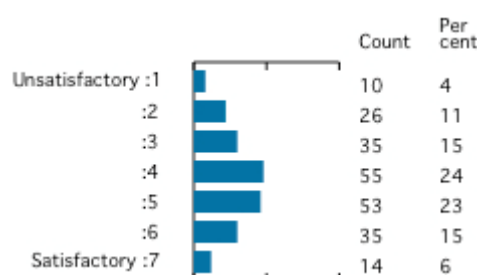


Figure 15: Distribution of scores (Building Use Studies, 2009)

The percentage of people dissatisfied has been a more frequent part of the recent reports, as it potentially allows for a comparison to other, more numerical methods of calculating which aim to have 80% of people at a level they consider comfortable. In the most recent reports, the percentage of people dissatisfied for each question is shown in the bar graph, stated for each question and then stated again on the summary page. The percentage of people satisfied will be different over the various attributes and the more variables that are looked at the less people will be satisfied with all of them. 'In practice, sensitivity to multiple factors is not perfectly correlated, and so as more factors of the office environment are considered, there will be fewer people satisfied with all of them' (Aronoff & Kaplan, 1995, p. 320).

The benchmark, the scale midpoint, and their corresponding confidence intervals form a critical region, explained below, which forms the basis of whether the building is rated as performing 'Better', 'Similar' or 'Worse.'

The slider displays the mean of the scores for the building compared to the mean and confidence intervals for the database buildings. The dashes on the top of the scale represent the benchmark mean and confidence limits, while the dashes on the bottom are the same for the scale midpoint. A critical region is made between the lowest and the highest of the confidence limits. The example below is an example of a building scoring similarly to the benchmark, as it is in the critical region.

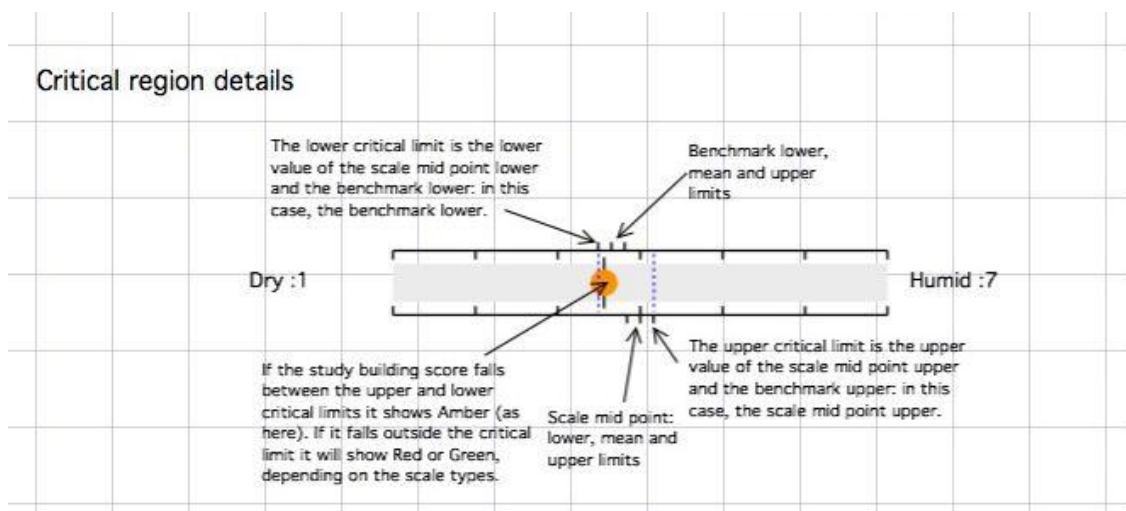


Figure 16: Critical region (Building Use Studies, 2009)

The percentile graph displays the same information as above but also shows where this building sits for this variable in relation to the other buildings in the database. The cross on the graph represents the midpoint of the scale. The blue dotted lines represent the upper and lower bounds of the critical region. For the example below the building is in the 31st percentile, and is within the critical region so is again rated as being similar. The group of buildings shown below is the NZ building database and consists of 24 buildings.

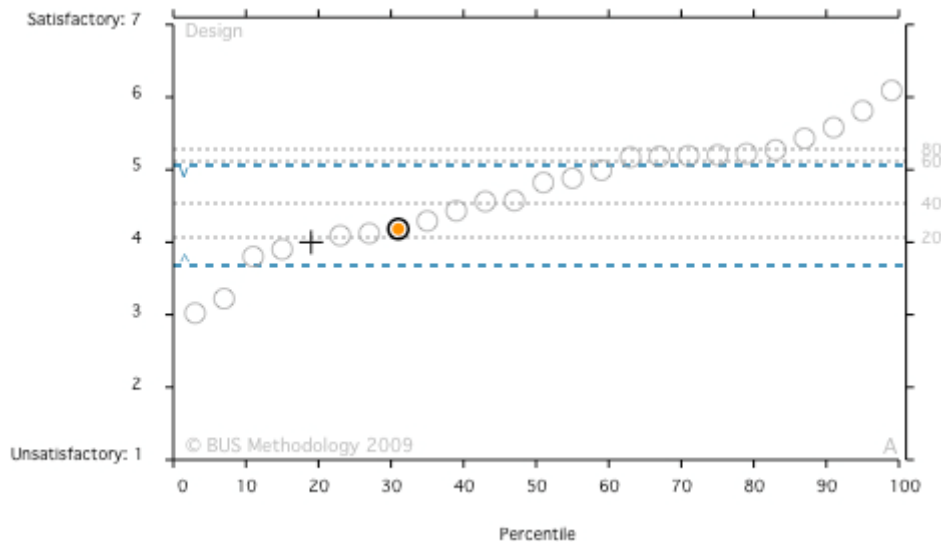


Figure 17: Percentile graph (Building Use Studies, 2009)

The quintiles are also displayed on this graph as horizontal grey dotted lines, as a numeric representation of the distribution. Quintiles are similar to quartiles, as the data is organised in ascending order and split into even groups. With quintiles it is groups of five rather than four. In this example, the building is performing in the second quintile of this group of buildings.

In the later reports that Building Use Studies produce more emphasis is placed on the quintiles, and they are referred to more often. The quintiles are stated for each question, as well as being displayed in the graphs.

3.4.1 Indices

In order to get an idea of the overall scoring of the occupants for the building some summary indices were created. These were developed during the Probe study. 'These indexes provide snapshots of how well a building works for its occupants, and are a first step in presenting results' (Leaman & Bordass, 2001, p. 130).

For each of these, 'z-scores' are first calculated. Z-scores are standard scores that are used to compare different distributions. These are comparable as they are standardised in units of standard deviations. They are calculated by dividing the amount that a score differs from the mean by the standard deviation (Salkind, 2005).

Equation 3: Standard score for BUS data

$$z = \frac{(X - \bar{X})}{s} \quad \text{in this situation} \quad z = \frac{\text{Actual} - \text{Benchmark}}{\text{Benchmark Standard Deviation}}$$

Three indices have been developed. The first is based on comfort. 'The *Comfort Index* attempts to encapsulate, in a single figure, an overview of occupants' perceptions of the building's comfort performance' (Baird, Christie, Ferris, Goguel, & Oosterhoff, 2008, p. 5). This takes into account seven variables; *Temperature Overall in Summer*, *Temperature Overall in Winter*, *Air Overall in Summer*, *Air Overall in Winter*, *Lighting Overall*, *Noise Overall*, and *Overall Comfort*. It looks at the average of 'z-scores' (Salkind, 2005). An example is shown below in Equation 4.

Equation 4: BUS Comfort Index (Building Use Studies, 2008)

$$\text{Average of z-scores of ...}$$

$$\text{BUS Comfort Index } 0.66 = \frac{\frac{\text{ComfOver} (4.94 - 4.41)}{0.65} + \frac{\text{LtOver} (5.26 - 4.85)}{0.46} + \frac{\text{NseOver} (4.30 - 4.24)}{0.66} + \frac{\text{TSoVer} (4.44 - 4.03)}{0.62} + \frac{\text{TWOVer} (4.59 - 4.21)}{0.57} + \frac{\text{AirSOVer} (4.28 - 3.97)}{0.55} + \frac{\text{AirWOVer} (4.50 - 4.05)}{0.48}}{7}$$

7

The second is the *Satisfaction Index* and uses the responses from four variables. ‘The variables defined as contributing to the users’ perception of satisfaction were designated to be design, needs, health and productivity’ (Baird, Christie, Ferris, Goguel, & Oosterhoff, 2008, p. 5). This is designed to do the same as the *Comfort Index* but for the satisfaction of the occupants.

Equation 5: BUS Satisfaction Index (Building Use Studies, 2008)

$$\text{BUS Satisfaction Index } 0.30 = \frac{\text{Average of z-scores of ...}}{4}$$

Design		Needs		Health		Productivity
(5.20 - 4.56)	+	(5.16 - 4.63)	+	(3.28 - 3.55)	+	(-1.57 - -2.69)
0.72		0.65		0.38		5.22

The resultant scores for both of these indices range from -3 to +3, from the 1-7 scales. This is due to the fact that they are now measures of standard deviation, similar to what was discussed in **Section 3.1: Nature and characteristics of benchmarks**. The higher the score the better the occupants are rating the building. The majority of the scores, approximately 95%, will lie between ± 2 . These scores are also portrayed as a percentile of where they fit in the dataset, similar to Figure 17.

A later addition was a combination of both of these and is referred to as the *Summary Index*. This is the arithmetical mean of both the comfort and satisfaction indices.

Equation 6: BUS Summary Index (Building Use Studies, 2008)

$$\text{BUS Summary Index } 0.48 = \frac{\text{Comfort index } 0.66 + \text{Satisfaction index } 0.30}{2} \dots \text{with equal weighting to comfort and satisfaction.}$$

The *Forgiveness Factor* is determined in the report of the results for a specific building by Building Use Studies. This factor aims to ‘quantify the tolerance occupants have for chronic faults’ in the building (Building Use Studies, 2008, p. 23). It is calculated by comparing the *Overall Comfort* score to the mean of the different variables. The *Forgiveness Factor* relates to indoor environmental quality only.

Equation 7: BUS Forgiveness Factor (Building Use Studies, 2008)

$$\text{Forgiveness } 1.08 = \frac{\text{ComfOver } 4.94}{\frac{\text{AirWOver } 4.50 + \text{AirSOver } 4.28 + \text{TWOVer } 4.59 + \text{TSoVer } 4.44 + \text{LtOver } 5.26 + \text{NseOver } 4.30}{6}}$$

These normally range from 0.8 to 1.2. The values above 1 indicate relatively more forgiveness (Building Use Studies, 2008).

These indices are also displayed in a percentile score in relation to the BUS buildings, similar to the graph described earlier (Figure 17). This shows where the particular building sits in relation to the BUS database buildings that it is being compared to.

The variables chosen to include in these indices were determined through Adrian Leaman’s years of experience. He stated that ‘comfort is self-evident. The *Satisfaction Index* variables are a selection of the ‘soft variables’ which are most commonly consulted. The *Forgiveness Index* is based on those variables that have overall scores as well as more detailed ratings’ (Leaman, 2010).

3.4.2 BUS rating scores

‘Two rating scales, one made up of ‘Ten Factors’ and one using ‘All Factors’ (around 45 in total), are used to provide an overall performance assessment of the buildings from the users’ point of view’ (Baird, 2010, p. 13).

The first is the ‘BUS rating score: selected variable method.’ This uses 10 variables including; *Overall Comfort, Design, Health, Image, Lighting Overall, Needs, Noise Overall, Perceived Productivity, Temperature Overall in Summer* and *Temperature Overall in Winter*. The ‘Better’, ‘Similar’, or ‘Worse’ rating given to the individual question with the comparison of the dataset is then looked at. A corresponding number is assigned depending on the rating. ‘Worse’: 1, ‘Similar’: 2, 3 or 4 and ‘Better’: 5. The ratings for ‘Similar’ are a judgement call depending on where the scores fit in the band. An example of this process is included in Appendix B. As the best possible score would be 50 (10 variables each having a ‘Better’ (5) rating), the actual score is portrayed as a fraction and then changed into a percentage value. This percentage is then transformed into a score out of 7 and this is the overall rating.

The scores are equivalent to:

- 1 (0-14.3%) Very poor
- 2 (14.4-28.6%) Poor
- 3 (28.7-42.9%) Below average
- 4 (43-57.2%) Average
- 5 (57.3-71.4%) Above average
- 6 (71.5-85.7%) Good practice
- 7 (85.8-100%) Exceptional

In this case the scores are split into equal bands that are not reliant on the building stock, i.e. they are not percentiles of a larger database as the majority of the energy benchmark examples were.

The second method, ‘BUS rating score: all variable method,’ is very much the same as the first, but takes into account all variables including the subset questions such as *Noise from Colleagues* and *Natural Light*. This produces an equivalent score. As this takes into account more variables the scores are generally not as high. This is shown in the 30 buildings analysed by Baird (2010) looking at both the BUS rating score: selected variable and the BUS rating score: all variable methods, ‘it is evident that taking all 45 or so factors into account proved to be a tougher test on the 7-point rating scale for most of the buildings’ (Baird, 2010, p. 14).

These last methods are only present in some of the reports as the benchmarks have developed over time. The most recent reports include the individual comparison, percentile rating, and the summary variables; although rather than displaying this as a number it is displayed as an indication of the number and percentage on a graph, similar to the percentile graph shown above. This is still calculated for these buildings and is available on request, but is not shown in the basic appendices.

The way that Building Use Studies benchmarks the buildings covers a wide range of different formats, from the individual questions through to a one number rating. It is difficult to determine which is the most suitable to use in New Zealand, as they all have their advantages. They seem to work best in conjunction with each other, each informing the other; the single number format perhaps is the easiest to understand for the client, but loses important information of the detail offered by the individual comparison.

3.5 Current building rating tools

A common application of benchmarks in the building industry is for use in building rating tools.

‘Assessment methods generally have recognisable ‘frameworks’ that organize or classify environmental performance criteria in a structured manner with assigned points or weighting’ (Cole R. J., 2005, p. 456).

There are many building rating tools that are used around the world to assess the sustainability of buildings. The large majority of these focus on the design stage of new buildings.

'In addition, there are many buildings in the United States, United Kingdom, and elsewhere that are claimed to be "green", "low energy", or "high-performance" buildings. It is never entirely clear on what evidence or data these claims are based. Such claims of high performance cannot be credible without standardised protocols that are applied consistently to the assessment of building performance' (ASHRAE; USGBC; CIBSE, 2009, p. 1).

Some of the main building rating tools have previously been focussed on the assessment of buildings at the design stage. Due to the inconsistency in claims for 'energy efficiency' etc., it is becoming more common for the rating tools to be applied to the actual performance of existing buildings. The main focus of most of these is on technical aspects such as the amount of energy and water that they are using, but some of them are also starting to consider how the buildings are working for the occupants.

This section will look at several building rating tools that are used currently, before focussing on those that already, or have plans to, incorporate the perceptions of people into these tools. It should be noted that the following list is by no means exhaustive of the different rating tools but are a selection of the most commonly used. They have been reviewed numerous times, such as Cole (2005).

3.5.1 Existing building rating tools

The selection of building rating tools that will be described further are:

- Green Star NZ – Office
- BREEAM
- LEED
- CASBEE

Green Star NZ – Office

Green Star NZ – Office Design Tool was largely based on Australia's equivalent, Green Star Australia and was first released in 2007. It assesses the buildings in terms of eight different aspects, which all feature different weighting; Management (10%), Indoor Environment Quality (20%), Energy (25%), Transport (10%), Water (10%), Materials (10%), Land Use and Ecology (10%) and Emissions (5%).

There are set criteria for each aspect and if the building reaches those criteria it receives a predetermined number of points. The eventual outcome is an overall star rating. Fixed ratings are used for a comparison. For example the building has to have a predicted energy use index of below 105kWh/m².yr in order to receive the credits for this point.

The indoor environmental quality section includes the ventilation rates, indoor air quality in terms of emissions from materials and air changes, thermal comfort and control, daylight and glare, external views, lighting levels and internal noise levels.

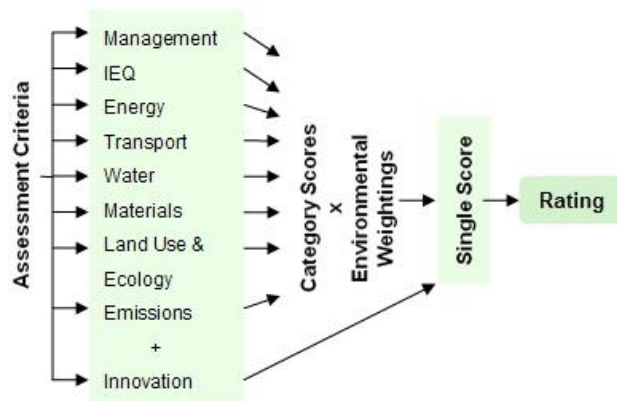


Figure 18: Basic structure of Green Star NZ (NZGBC, 2008)

The score out of 100 that the building can receive then corresponds to a star rating that represents the building's level of performance. The maximum number of stars a building can receive is six. The table below shows the relationship between the number of stars and the level of performance (NZGBC, 2008).

Table 4: Green Star rating tool scores

One Star	Score of 10	Represents Minimum Practice
Two Stars	Score of 20	Represents Average Practice
Three Stars	Score of 30	Represents Good Practice
Four Stars	Score of 45	Represents Best Practice
Five Stars	Score of 60	Represents NZ Excellence
Six Stars	Score of 75	Represents World Excellence

BREEAM

Building Research Establishment Environmental Assessment Methodology (BREEAM) is the equivalent rating system for the UK; it was introduced in 1990 and was the first of its kind. It was developed by the Building Research Establishment (BRE) as a voluntary assessment method of evaluating a building and producing a single score rating.

Similar to the other rating tools the methodology is made up of different sections each with their own weighting. For this method they are Management (12%), Health and Wellbeing (15%), Energy (19%), Transport (8%), Water (6%), Materials (12.5%), Waste (7.5%), Land Use and Ecology (10%), Pollution (10%), and an additional 10% for Innovation.

The resultant percentage score, if above 30%, is then classified into five rating categories from a Pass of 30-35% to an Outstanding of 85% or more.

It is one of the more widely used assessment methods and is able to be used internationally with BREEAM – International (BRE, 2010).

LEED

Leadership in Energy and Environment Design (LEED) was first introduced to the market in 1998, and is the equivalent building rating system used in USA. It is based largely on other methods, specifically BREEAM. LEED was developed largely to define green buildings by establishing a common standard of measurement, promote integrated design, recognise environmental leadership in the building industry, create green competition, raise awareness of green building benefits and transform the building market (Bailey, 2010).

'LEED points are awarded on a 100-point scale, and credits are weighted to reflect their potential environmental impacts' (USGBC, 2011).

There are also an additional 10 points awarded for innovation. The credits are achieved from nine areas, including the one for innovation. These are: Sustainable sites, Water efficiency, Energy and atmosphere, Materials and resources, Indoor environmental quality, Locations and linkages, Awareness and education, Innovation in design, and Regional priority. At present there are nine different rating systems used for different building types. The weightings applied to the credits from each area depend on the rating system being used. Examples of the distribution of credits for 'New Construction' and 'Existing Buildings: Operations & Maintenance' are displayed below in Figure 19 and 20.

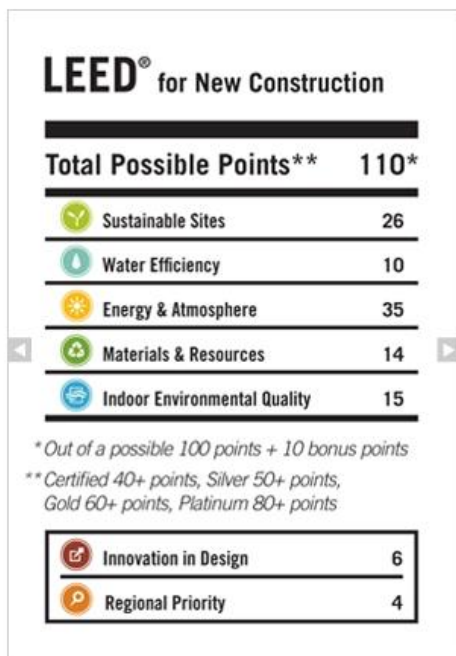


Figure 19: New Construction (USGBC, 2011)

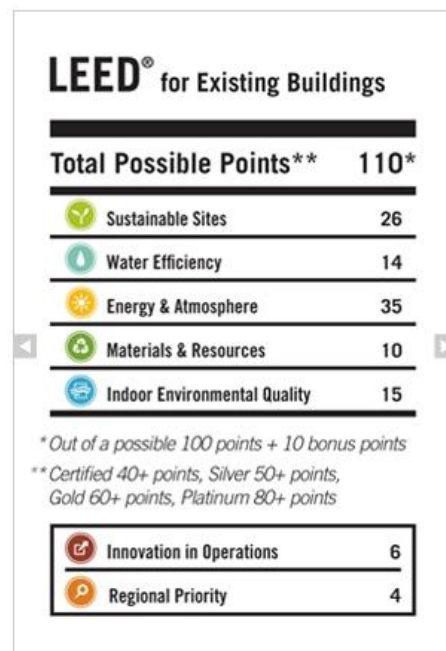


Figure 20: Existing Buildings (USGBC, 2011)

Once the building has been evaluated in terms of the correct credits and weightings it is awarded an overall rating. These are dependent on the total number of credits achieved. Certified (40-49 points), Silver (50-59 points), Gold (60-79 points), Platinum (80 points and above) (USGBC, 2011).

CASBEE

The Comprehensive Assessment System for Building Environmental Efficiency (CASBEE) is a building assessment method in Japan. It is fairly new and was introduced in 2002.

'CASBEE: while employing an additive/weighting approach, it breaks away from the simple addition of points achieved in all performance areas to derive an overall building score, which has been the dominant feature of virtually all previous methods' (Cole R. J., 2005, p. 459).

The focus of CASBEE is a ratio of Environmental Loading (resource use and ecological impacts) against Environmental Quality and Performance (indoor environmental quality and amenities). The results that are produced are then providing the client with an assessment of the environmental implications associated with the specific level of service that the building environment provides for the occupants. The results are displayed graphically with the Environmental Loading plotted against the Environmental Quality and Performance and the resultant value displayed as the gradient of the lines connection the results and the origin points. These results can then be presented into five classes which are compatible with five star ratings. The classes are C (less sustainable), B-, B+, A and S (more sustainable) (Boonstra & Pettersen, 2003). An example is displayed below in Figure 21.

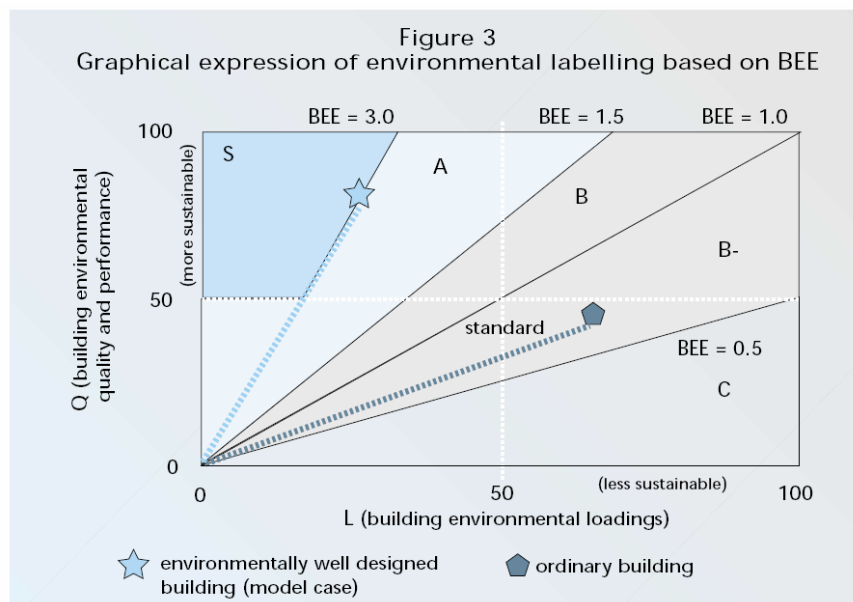


Figure 21: CASBEE example (Boonstra & Pettersen, 2003, p. 82)

3.5.2 Comparison of the indoor environmental quality category

These Building Sustainable Rating Tools (BSRTs) consist of many areas of building performance, each with a different weighting. Jonathan Smith, as part of his studies at Victoria University assessed and compared the different building sustainability rating systems that could be appropriate to use in New Zealand. The NZ Tool referred to a set of draft weightings for aspects of sustainability in NZ determined from a survey of building industry BSRT experts (Smith, 2008).

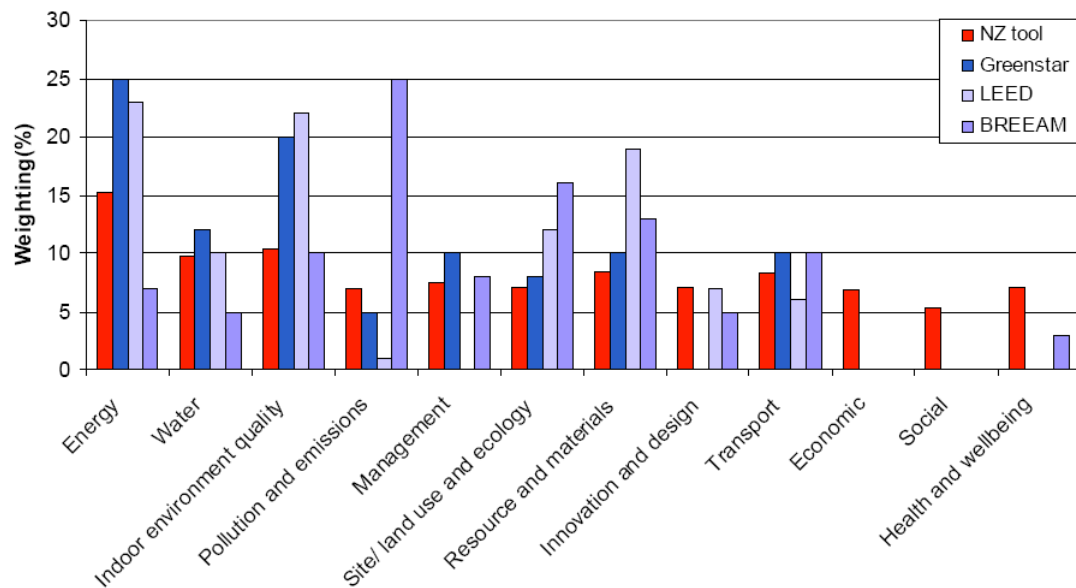


Figure 22: Comparison of environmental aspects weightings (Smith, 2008, p. 43)

The categories that relate to the experience of the occupants in the building are ‘Indoor environment quality’ and ‘Health and wellbeing.’ In terms of indoor environmental quality the building rating tool with the highest weighting on this is LEED, second only behind its weighting of energy. The BSRTs experts rated an additional 7% in relation to the health and wellbeing of the occupants in the NZ tool.

This corresponds with additional findings in the comparison of LEED and BREEAM where:

‘The LDY Eco Systems team attempted to fit the LEED credits into the BREEAM UK categories, such as water, energy, pollution, air quality, ecology, use of land and transport, and found that LEED gives slightly more importance to the occupant’s health and comfort, while BREEAM UK and BeSpoke Checklists would tend to be more focused around environmental impacts’ (Julien, 2009, p. 31).

In New Zealand when the NZGBC was initially developing the Green Star NZ tool, they held a number of industry workshops where over 200 industry professionals discussed the makeup of a building rating tool specific for New Zealand and the importance of the different aspects. They were asked, in groups, to rank from 1 to 10 typical environmental aspects and their categories.

Table 5: Rating tool workshop environmental aspect weighting (NZGBC, 2006, p. 25)

Topics measured by the tools we are looking at	Overall	AKL	WTG	CHC
Energy efficiency	1	1	1	1
Environmental quality within the building	2	2	2	5
Transport	3	3	6	8
Materials	4	5	6	6
Management systems, procedures and plans	5	6	4	3
Land use and ecology	6	4	3	3
Water efficiency	7	8	9	9
Flexibility and adaptability	8	9	5	2
Emissions, effluent and pollution	9	7	8	6
Quality, service and risk	10	10	10	10

Overall the industry professionals in NZ rated the importance of the environmental quality within the building as being second only to energy efficiency. This was consistent from the groups in Auckland and Wellington and it is only in Christchurch where the rating was lower.

This reinforces the importance, in the minds of industry professionals, of the environmental quality of a building as a very important part of a building assessment tool. The perceptions of the users and how the indoor environment affects them, which is the main focus of this research report, fits mainly into this important section in terms of building performance, and several building rating tools are beginning to incorporate these.

‘A consistent feature of these developments is that most discussions are technically framed, i.e., emphasising technical systems and their attendant potential reduction in resource use and resultant ecological loadings. This focus makes scant reference to how users interact with technical features and systems or, more generally, to the cultural acceptance of green buildings. In practice, however, a building designed with excellent ‘green’ performance standards can be severely compromised because the specification and technical performance fail adequately to account for the inhabitants’ needs, expectations and behaviour. Moreover, long-term, broadly based solutions to environmental problems will depend on major changes in human values and actions’ (Cole R. J., 2003, p. 57).

3.5.2 Current building rating tools that incorporate perceptions

National Australian Built Environment Rating System (NABERS)

NABERS is a voluntary rating that can be initiated by a building owner, manager or tenant. There are three types of rating systems that can be carried out; ‘Base building’ looks at the building by itself with no occupants in it; ‘Whole building’ looks at the entire building with all tenants in; and ‘Individual tenancies’ looks at specific tenancy areas. It is aimed specifically at buildings in operation. It reports separately the ratings for four categories: Energy, Water, Waste and Indoor Environment.

The aim of the Indoor Environment rating is that it ‘provides market recognition and a competitive advantage for buildings with a high quality, healthy and environmentally friendly indoor environment,’ and ‘encourages best practice in indoor environment practices of commercial buildings to minimise the building’s impact on the occupants’ (Australian Department of Environment, Climate Change and Water, 2010). It bases its scoring on both physical measurements and an occupant questionnaire.

The indoor environment rating is split into five different aspects, each with different number of points. The questionnaire contributes all or some of the scoring of these and the outcome of the survey is a ‘building scorecard that lists the percent satisfaction for various indoor environment categories (thermal comfort, air quality, acoustic comfort, lighting quality and office layout). For each indoor environment category this percent satisfaction is applied to the total allocated points of the respective category to determine the actual points attained in that category’ (Australian Department of Environment, Climate Change and Water, 2010).

Table 6: NABERS protocol (Baird, 2009, p. 1073)

Parameters and Relative Weightings (Whole Building only)	Whole Building	Tenancy only	Base Building
Thermal Comfort (30%)	P (15pts) + S (15pts)		P (15pts)
Air Quality (20%)	P (15pts) + S (15pts)	P (15pts) + S (15pts)	P (15pts)
Acoustic Comfort (20%)	P (15pts) + S (15pts)	P (15pts) + S (15pts)	P (15pts)
Lighting (10%)	P (15pts) + S (15pts)	P (15pts) + S (15pts)	
Office Layout (20%)	S (30pts)	S (30pts)	

P denotes physical measurements; S denotes survey questionnaire; the numbers in brackets (15pts), (30pts), etc., indicate the maximum number of points that can be awarded for each aspect.

Green Star NZ - Performance

Currently the New Zealand Green Building Council (NZGBC) is working in conjunction with the Energy Efficiency Conservation Authority (EECA) to create a performance rating tool, based largely on the methodology from the NABERS.

The existing NZ Green Star Office rating tool has sections that were previously described. Some of these will be able to have corresponding performance ratings, but not all of them. 'Energy' and 'Water' are sections where performance in practice can be tested by carrying out an audit and determining how much energy/water the building is actually using. 'Materials' is more focussed on the actual building phase rather than performance, and similarly with 'Land Use and Ecology.' 'Emissions' can also be looked at in performance, as can 'Indoor Environment Quality' and 'Management.' 'Transport' aims to encourage the use of less harmful ways of travelling. The rating of buildings by its users that is the focus of this study could be incorporated in the section in 'Indoor Environment Quality.' The current proposal is a combination of physical measurements and occupant questionnaires, similar to NABERS.

LEED – Existing Buildings: Operations & Maintenance and New Building (EBOM)

LEED is also beginning to use occupant questionnaires in their rating tool of existing buildings. In LEED EBOM IEQ Credit 2.1 is labelled 'Occupant Comfort – Occupant Survey'. The intent of this credit is 'to provide for the assessment of building occupants' comfort as it relates to thermal comfort, acoustics, indoor air quality (IAQ), lighting levels, building cleanliness and any other comfort issues' (USGBC, 2011, p. 68). This credit requires a survey response rate of 30%, and a complaint response system to be in place. The general requirement for this is a percentage of people satisfied with the building of 80% (USGBC, 2009, p. 76). Although it does not specify the specific questionnaire to use, CBE has a form which will allow you to assess your building and apply for the credits.

LEED also has another credit including a thermal comfort survey in their new construction rating tool. In this credit the owners have to agree to undertake a thermal comfort survey 6-18 months after the building has been tenanted, and agree to plan for corrective action if more than 20% of the occupants are dissatisfied with the thermal comfort in the building. It is a statement of intent rather than an analysis based on results (USGBC, 2009, p. 76).

HOPE project

Although the HOPE project is not an official building rating tool, it includes a unique way of transforming results from an occupant questionnaire to an overall building rating, taking into account both qualitative and quantitative data. The HOPE project has been briefly described in **Section 2.3: Examples of questionnaires used**. In the case of HOPE, it rates the scores for each aspect as 'green' 'red' 'yellow' or a 'black' class.

The HOPE project questionnaire is made up of questions about the Health and Comfort of the occupants in the buildings, and a simultaneous energy audit is completed. The Health, Comfort and Energy Use are then looked at individually for the sorting, due to the different type of data they all produce. The comfort questions are all asked on a 7-point scale with the majority ranging from 1: Satisfactory to 7: Unsatisfactory, with some with an ideal score of four. For each of these aspects the buildings are sorted into mainly two classes, with 'poor' 'unsatisfactory' or 'red' on one side, and 'good' 'satisfactory' or 'green' on the other. If it is not clear where the building fits then it is classified as 'yellow' or 'uncertain' and if the results are at a very low level that cannot be balanced out by a particularly good score in another category it is classified as a 'black' or 'veto' rating.

The bilateral scale questions are transformed to a -3, 0, +3 scale by subtracting the average score, where four is the ideal, by four (Roulet, Flourentzou, Foradini, Bluyssen, Cox, & Aizlewood, 2006).

Table 7: HOPE classifications (Roulet, Flourentzou, Foradini, Bluysen, Cox, & Aizlewood, 2006)

Comfort scores from 1-7 scale	Classification
<2.5	Satisfactory/Green
≥4	Unsatisfactory/Red
>6	Veto/Black
Comfort scores from -3-+3 scale	Classification
-0.75-+0.75	Satisfactory/Green
Outside ±0.75	Unsatisfactory/Red
Outside ±2.5	Veto/Black
Health (BSI) quantiles	Classification
< 35%	Satisfactory/Green
≥70%	Unsatisfactory/Red
Outside 2 SD	Veto/Black
Energy	Classification
<150kWh/m ²	Satisfactory/Green
>250kWh/m ²	Unsatisfactory/Red
>500kWh/m ²	Veto/Black

The overall building is then considered satisfactory/green if there is a majority (50% or more) of criteria with green marks and no veto. The building is unsatisfactory/red if there are more than 50% of the criteria with red mark, or fewer than 50% green marks or 33 red marks, or one veto. If the percentage of veto marks is larger than 33% the building is marked black. A building is considered yellow, or not sorted, otherwise (Roulet, Flourentzou, Foradini, Bluysen, Cox, & Aizlewood, 2006).

3.6 Benchmark summary

Benchmarks are simply a comparison point; an indication of whether something is rating better or worse than an average. This chapter has explored benchmarks of different types, starting from the statistical basis through to common applications. There are simple statistical methods that the majority of the examples use.

Through the assessment of the basic statistics and the existing methods the importance of the distribution became apparent. This will need to be considered in the future benchmarking analysis.

The complication in benchmarking in terms of occupant questionnaire data is the summary of the different variables that affect the indoor environment; and there are a variety of methods that are able to be used.

The most common measure of relative standing that was used in energy benchmarks, questionnaire examples and the rating tools were percentiles. These were either in equally spaced groupings, such as quartiles, or uneven groupings determined by the researcher.

The questionnaire examples represent the results of a building in two main ways; the percentage of people satisfied/dissatisfied, or the mean scores. These previous studies have determined that the rating scales used in the questionnaires can be considered interval scales, provided the 'verbal descriptor for each number is carefully chosen to enhance the perception of equal distance between the numbers' (Juniper, 2009, p. 1015). The rating tools that currently incorporate user perceptions have done so by using the percentage of people satisfied as a more comparable measure than a specific score.

The common application of benchmarks, building rating tools, employs different ways of combining several different aspects of a building to obtain an overall rating. The indoor environment is an important, and consistent, part of the different rating tools. Results from occupant questionnaires are able to be incorporated into these tools as a way of showing how the buildings are performing for the occupants. These rating tools include separate credits for different areas of performance, so often look at the main features that make up the indoor environment separately, e.g. Air quality, temperature, light, acoustics.

Context is an important issue in terms of benchmarks, and in the assessment of buildings. This emphasises the relevance and importance of country-based evaluations, through an attempt to compare factors as closely as possible using common criteria. It seems much more relevant to compare buildings from the same country which will have similar limitations. This is also followed through with an argument from Preiser (2010):

'Cultural differences need to be recognised and dealt with when carrying out POES on the same building types in different countries and cultural contexts' (Preiser, 2010, p. 458).

4.0 RESEARCH METHODS

In order to create a clear picture of what data the Building Use Studies questionnaire produces, and how potential approaches for the creation of benchmarks will be explored, this chapter is separated into four sections. Using the previous research into questionnaires and benchmarks, along with the specific data received for the pilot database from the BUS questionnaire, conclusions will be able to be made as to the appropriateness of the data, and potential ways of determining benchmarks and their incorporation into building rating tools.

- Section 4.1 describes the questionnaire that is used to survey the buildings, and the method used for distribution and collection.
- Section 4.2 provides a description of the buildings from which the database is compiled.
- Section 4.3 explores the methods used to analyse the data received from the pilot survey to determine patterns and influencing factors.
- Section 4.4 considers the method for exploring potential approaches for the creation of benchmarks for users' perception scores and the incorporation of these scores into rating tools.

The aim is to create a clear picture of what the questionnaire data gives, and how it will be analysed to get the information that will prove/disprove the hypothesis.

'Given a representative sample of building user questionnaires it should be feasible to develop user perception benchmarks for New Zealand buildings and incorporate them into BSRTs.'

4.1 BUS questionnaire design

Post Occupancy Evaluations aim to help with future planning and design of buildings. The particular POE questionnaire used was originally developed for the Probe (Post-Occupancy Review of Buildings and their Engineering) study, by Building Use Studies (Building Use Studies, 2009). It stemmed from the need for a way of assessing building performance when 'sick building syndrome' was a common issue. Originally it was a 16 page document but has evolved down to its current three page A4 single-sided version.

It is made up of four main parts:

Background – gives an idea of who the response is coming from, where the occupant is situated in the building, and how often they use it. This is made up of 13 questions in total.

Overall building – looks at the design, and how well the building as a whole and individual workspaces suit occupants' needs etc. There are a total of 11 rating scales and seven spaces for comments.

Indoor Environment – this part elicits responses relating to occupant thermal comfort, noise, lighting, overall comfort, productivity, health, personal control, response to problems and effect on behaviour. This is the largest section and consists of 39 rating scale questions and seven spaces for comments.

Travel to Work – seeks information on how long it takes and what method the occupants use. This is made up of eight questions and one space for general comments.

The questionnaire mainly uses a 7-point scale where occupants can rate how they find particular aspects of a building. One 9-point scale is used to assess *Perceived Productivity*, and small spaces are allowed where occupants can insert comments on 15 aspects.

The 7-point scale allows enough options to get an idea of the occupant's perceptions, as opposed to a 3-point scale which would give limited results. The odd number of points allow for a neutral status to be selected.

The scale questions themselves are considered to be ‘semantic differential’ scales. These consist of the scales having labels on either end, such as ‘unsatisfactory’/‘satisfactory’, or ‘uncomfortable’/‘comfortable’, with numbers in between allowing for interpretation by the occupant.

These scales are then further categorised into three types depending on the ideal score. These are categorised into ‘A’, ‘B’ and ‘C’ scales. ‘A’ scales have 1 as the worst and 7 as the ideal score. ‘B’ scales have an ideal score at the midpoint of 4, resulting in both 1 and 7 being lowest ratings. ‘C’ scales are the opposite of ‘A’ scales; the lower the score the better, where 1 is the ideal score and 7 is the worst rating. As the questions cover a wide range of topics, the variables that they are dealing with also differ. There are several ‘summary’ variables in the questionnaire. These are when the occupants are asked to rate the building overall in terms of specific aspects. For example, the noise section consists of one ‘summary’ variable dealing with *Noise Overall* before going into more detail with noise sources and types of problematic noise. All of the summary questions are ‘A’ scales, in most cases ranging from ‘unsatisfactory’ to ‘satisfactory’ with an ideal score of 7. When dealing with more specific sources in which there could be considered too much or not enough there is an ideal score of 4 (B scales) and for a small number there is an ideal score of 1 (C scales), such as *Unwanted Interruptions*.

The scale questions allow for a more quantitative analysis to be carried out while the additional comments (if respondents avail themselves of the opportunity) are more qualitative in nature. It is also quick and easy to complete. Experience has indicated that the occupants may be more willing to complete it because of this, and even if they do not leave any comments, an overall idea of how the building is working in practice will be obtained.

An example of the different types of questions that are in the questionnaire is shown below. This is the ‘noise’ section of the questionnaire and consists of the three types of questions and space for comments. They all consist of a 7-point scale with the ‘best’ score changing depending on the question. The ideal scores are highlighted in Figure 23.

Noise How would you describe noise in your normal work area?
This question refers to conditions all year round
Please tick your rating on each scale

Noise overall	Unsatisfactory	1	2	3	4	5	6	7	Satisfactory
Noise from colleagues	Too little	1	2	3	4	5	6	7	Too much
Noise from other people	Too little	1	2	3	4	5	6	7	Too much
Other noise from inside	Too little	1	2	3	4	5	6	7	Too much
Noise from outside	Too little	1	2	3	4	5	6	7	Too much
Please estimate how you are affected by unwanted interruptions ...									
Unwanted interruptions	Not at all	1	2	3	4	5	6	7	Very frequently

Comments about noise and its sources

Figure 23: Noise section of POE questionnaire

The Building Use Studies survey methodology (Building Use Studies, 2009) (‘Probe Approach’) was followed for these surveys. The paper-based questionnaires were handed out personally, ‘because these reach everyone in the building (including staff who do not have computers) and produce much higher response rates than internet-based questionnaires’ (Leaman, 2005, p. 24). This enabled an explanation to the occupants of what it was for, what they have to do and answer any questions that they have. It was a census sample, handing out to everyone in the building, with the aim of a response rate of at least 80%. The questionnaires were then personally collected over a number of days, to try and obtain the highest possible response rate.

Once the questionnaires were collected the data was input into a specially designed Excel spreadsheet and sent to the questionnaire creator, Adrian Leaman, who produced a summary report of the results. The general format and method of displaying results having been discussed previously in **Section 3.4: Assessment of BUS benchmarks**.

4.2 Pilot database

The main control on the buildings chosen for assessment in this research project is that they have to be situated in New Zealand, and are either commercial or institutional in use. They have been sourced from various places, to try and obtain a representative sample of the buildings throughout NZ.

The resultant database consists of 30 buildings sourced from:

- 28 previously surveyed buildings – Baird, VUW, E Cubed Building Workshop
- Two buildings surveyed during this study

These buildings were surveyed over a range of years from 1997-2010. The database could be increased in future by surveying buildings identified as part of the BEES project.

30 responses is generally considered the minimum sample size for making population estimates (Salkind, 2005). This database will allow flexibility in the statistical analysis, and exploration of how possible benchmarks could be constructed.

The resulting margin of error for this sample of 30 buildings, from a building population of 6,536 (Isaacs, et al., 2009), with a confidence level of 95% is $\pm 18\%$. The margin of error is equivalent to the values above and below the sample statistic (refer to Appendix C).

4.2.1 Previously surveyed buildings

Over the past nine years Post Occupancy Evaluations (POE) of buildings from around the world have been conducted using the BUS questionnaire, to test performance in practice through the occupants' perceptions of the indoor environment (Baird, 2010). These buildings are located in 11 different countries. The three New Zealand building results from this analysis are used towards the creation of this pilot database. All of these buildings are either commercial or institutional buildings.

'These were selected on the basis of their sustainability 'credentials'. Virtually all of them were recipients of national awards for sustainable or low energy design, or were highly rated in terms of their respective country's building sustainability rating tool or in some other way pioneered green architecture' (Baird, 2010).

They employ different sustainable features, such as different ventilation methods, layouts and materials.

Surveys of 12 New Zealand buildings were carried out by students under the supervision of Professor George Baird as part of their Masters and Honours programmes at VUW SoA from 2001 to 2009. The majority of the buildings are situated in Wellington with a few scattered around the country.

An additional 13 building surveys were carried out by E Cubed Building Workshop during the past few years and permission sought to use their results in the database. These had been commissioned by building owners, so although they may be a range of buildings there is the thinking that those buildings commissioned are often towards the higher end of performance, 'as managers who commission surveys are themselves motivated to improve building performance and are more likely to have good buildings' (Leaman, 1997, p. 38). In fact this was not such an issue in this sample, as not all of the surveys were carried out as Post Occupancy Evaluations. A number of the buildings surveyed by E Cubed Building Workshop were surveyed in preparation for a move, so were actually Pre-Occupancy Evaluations.

Table 8: POE details of the previously surveyed buildings

	Location	Building Use	Type of Respondents	Full/ Partial	Number of surveys	Year Surveyed	Who did the survey
Building 1	Wellington	Commercial	Staff	Full	196	2008	VUW
Building 2	Wellington	Commercial	Staff	Full	130	2008	VUW
Building 3	Auckland	Institutional	Staff	Full	25	2003	Baird
Building 4	Auckland	Commercial	Staff	Full	59	2005	Baird
Building 5	Dunedin	Institutional	Staff	Full	36	2004	VUW
Building 6	Palmerston North	Institutional	Staff	Full	86	2002	VUW
			Student	Partial	226		VUW
Building 7	Nelson	Institutional	Staff	Full	46	2001	VUW
			Visitor	Partial	144		VUW
Building 8	Wellington	Institutional	Staff	Full	11	2004	VUW
			Visitor	Partial	115		VUW
Building 9	Christchurch	Institutional	Staff	Full	57	2001	Baird
			Student	Partial	205		
Building 10	Wellington	Commercial	Staff	Full	106	2003	VUW
Building 11	Wellington	Institutional	Staff	Full	68	2002	VUW
			Student	Partial	325		VUW
Building 12	Wellington	Commercial	Staff	Full	233	2009	VUW
Building 13	Wellington	Commercial	Staff	Full	341	2009	VUW
Building 14	Wellington	Commercial	Staff	Full	102	2009	VUW
Building 15	Wellington	Commercial	Staff	Full	123	2009	VUW
Building 16	Auckland	Commercial	Staff	Full	53	2006	E ³ BW
Building 17	Auckland	Commercial	Staff	Full	63	2006	E ³ BW
Building 18	Auckland	Commercial	Staff	Full	29	2006	E ³ BW
Building 19	Auckland	Commercial	Staff	Full	35	2006	E ³ BW
Building 20	Wellington	Commercial	Staff	Full	51	2006	E ³ BW
Building 21	Wellington	Commercial	Staff	Full	32	2006	E ³ BW
Building 22	Wellington	Commercial	Staff	Full	80	2006	E ³ BW
Building 23	Wellington	Commercial	Staff	Full	137	2006	E ³ BW
Building 24	Wellington	Commercial	Staff	Full	94	2005	E ³ BW
Building 25	Wellington	Commercial	Staff	Full	111	2008	E ³ BW
Building 26	Auckland	Commercial	Staff	Full	112	2009	E ³ BW
Building 27	Auckland	Commercial	Staff	Full	40	2009	E ³ BW
Building 28	Auckland	Commercial	Staff	Full	439	2007	E ³ BW

4.2.2 Buildings surveyed during this study

Two buildings had Post Occupancy Evaluations carried out on them by the author. Both of these buildings had sustainable design intentions, one being awarded 5 stars in the Green Star NZ system. One is situated in Wellington, and the other in Waitakere, Auckland.

'Aorangi House is a 12-storey office building on the periphery of Wellington's CBD that has recently been refurbished' (Marriage, 2010, p. 1). The refurbishment aimed to produce a design of 5-star green rating, and it has recently been awarded this. The ventilation system is mixed mode, with both manual and automated opening windows around the perimeter and an air conditioning system which is used only when the natural ventilation is insufficient. The lower 10 floors of the building are occupied by the engineering firm BECA, and the top two are occupied by the NZ Correspondence School, Te Kura.

Central One, Waitakere Central is a 3-storey building situated in Waitakere, Auckland. It is multi-tenanted; at the time the survey was undertaken there were 12 different firms, with three retail spaces on the ground floor vacant. The top two floors are solely made up of offices, while the ground is a mix of office and retail, including a large cafe. The brief for Central One was to provide 'A-grade, urban design-focused commercial office and retail space that embraces the principles of Environmentally Sustainable Design (ESD)' (Waitakere Properties Limited, 2007, p. 1). It was designed to be of 4-star green star rating performance but is yet to be assessed for actual performance. It has full air-conditioning and sealed windows.

Table 9: POE details of the buildings surveyed during this study

	Location	Building Use	Type of Respondents	Full/ Partial	Number of surveys	Year Surveyed	Who did the survey
Building 29	Wellington	Commercial	Staff	Full	192	2010	Author
Building 30	Auckland	Commercial	Staff	Full	64	2010	Author

4.2.3 Buildings that could be surveyed in the future - sourced from BEES project

As it is desirable for the sample of buildings to be larger in size and as random as possible, the main building source that could eventually be applied to the benchmark methodology in NZ is through the Building Energy End-Use Study (BEES). The BEES Project is focussed on developing building energy and water performance benchmarks specific to NZ. It therefore needs to be based on a representative sample of the building stock. These buildings are not user surveyed at present.

A main focus of the BEES project was to gain a representative sample of buildings to assess their energy and water usage. The BEES project is adopting a three-level survey approach – Aggregate (a large number of randomly selected buildings), Targeted (for around 300 buildings) and Case studies (of a small number). The hope is that it will be feasible to undertake user surveys of the Targeted and Case study samples; therefore the data from these could be used for further application of this research allowing a profile of the existing building stock in regards to user perceptions to be determined, as well as establishing representative benchmarks.

From the BEES analysis of the NZ building stock, it was determined that there were about 6,536 office buildings throughout the country (Isaacs, et al., 2009). Therefore to achieve a confidence level of 95% and a margin of error of $\pm 5\%$, the survey will require a sample of 363 buildings to ensure a sample that reflects the sample population (refer to Appendix C) (Creative Research Systems, 2011), however a sampling frame will still need to be used to determine that the buildings are representative.

4.3 Analysis of scores

The Post Occupancy Evaluations of the buildings are put in a database so the results can be compared to each other. This database is made up of the mean scores entered into an SPSS spreadsheet for further analysis. SPSS is an advanced statistical analysis program.

The questions that feature in the BUS questionnaire are 'semantic differential' and previous studies have determined that these can be considered interval scales and thus the mean is the appropriate measure of central tendency to use.

'With careful construction, interval response scales provide data that the majority of statisticians agree meet the assumptions for parametric analysis' (Juniper, 2009, p. 1015).

This methodology section focuses on the questionnaire data that have been received and how it relates to itself and the buildings it has come from. There are a large number of factors that need to be taken into account in the development of benchmarks. In this research, they have been narrowed down into five main steps.

- Classification of design
- Benchmark comparisons
- Questionnaire scores
- Questionnaire variables
- Building variables

4.3.1 Classification of design

It was determined in **Chapter 3.0: Benchmarks** that it is necessary to consider the buildings that the results are from in order to ensure the benchmarks that are produced are appropriate. The buildings in the pilot database are compared in terms of their design, use, year occupied, and location. This enables an overall assessment of the buildings' design features to gain an understanding of the representative nature of the sample, and to determine influential characteristics.

A table is produced to classify the designs of the buildings. This allows for easier comparison between design and for grouping similar features together. Classifying the buildings in this way provides an overall idea of the different designs, and an understanding of the buildings themselves before analysing how they are performing. The different designs, when applicable, are compared to the BEES data and sampling method, as in the BEES project some main proportions in the current NZ building industry were determined.

In the BEES project, to ensure that the sample obtained was representative, the building stock in NZ was split into different groups.

- Five floor area groups, each with about 9.8 million m² floor area – under 650m², under 1,500m², under 3,500m², under 9,000m² and over 9,000m².
- Five use groups – office, retail, mixed, service and warehouse.
- Two geographical groups – Auckland and the rest of New Zealand (Isaacs, 2010, p. 70).

Out of the five use groups, for this analysis only the office buildings are used to determine the benchmarks for office and institutional buildings in New Zealand. As ideally there would be separate benchmarks for each use group, this could be explored at a later stage.

While the majority of these classifications are straightforward or easily understood, the ‘design intent’ is slightly more flexible as the term ‘sustainable’ can have different meanings. In terms of this report and the pilot database, the buildings were referred to as ‘sustainable’ where they had been created with explicit intent to be low energy users and to include environmentally sustainable design features and principles. The classification into a sustainable set of buildings was carried out similarly to previous studies (Leaman, Thomas, & Vandenberg, 2007; Leaman & Bordass, 2007; Baird & Oosterhoff, 2008; Baird, Christie, Ferris, Goguel, & Oosterhoff, 2008; Baird, 2010). Whether in practice this is followed through to a building that has less environmental impact is not tested. All other buildings were classed as ‘conventional’.

These classifications are referred to when analysing the scores to get a further understanding and see if there was any correlation between building design and the occupants’ POE scores. A summary of the design features identified is displayed in Table 10.

Table 10: Classification of design categories

Building Features	Options
Location	City in New Zealand
Building Use	Commercial, Institutional
Building Age	<1970, 1970-2000, >2000
Building Height	<3, 3-6, >6
BEES Floor Area Grouping	1, 2, 3, 4, 5
Ventilation Method	NV, MM, AC
Design Intent	Sustainable, Conventional

4.3.2 Benchmark comparison

Using the most recent data of the BUS benchmark, UK and NZ versions, comparisons are made to see how buildings scored and differences in the overall ratings. A simple comparison of mean scores and 95% confidence levels of the summary variables are carried out, followed by a more inclusive analysis of all the questions in the database, and ranking of the best score out of the databases for all the questions.

The means and confidence intervals are used as they correspond with the methods BUS use in their report; they are then comparable with the BUS benchmarks.

The most recent of the databases from this sample are the BUS UK 2006 benchmark and the BUS NZ 2010 benchmark. The BUS NZ benchmark uses the results from the last 24 buildings surveyed in NZ. These are likely to be made up of the VUW SoA and E Cubed Building workshop surveys described above given the small scale of such work carried out in NZ. Survey confidentiality means that this cannot be confirmed. The BUS UK 2006 benchmark is compiled of the 50 buildings in the UK that were most recently surveyed at the time of the benchmark.

In accordance with the BUS methodology, the buildings are compared to the databases (both the UK 2006, and the NZ 2010 separately) and rated as ‘Better’, ‘Similar’ or ‘Worse’ for each question. As described in **Section 3.4: Assessment of BUS benchmarks**, they are considered ‘Better’ or ‘Worse’ when outside the 95% confidence levels of the database and ‘Similar’ when within the range. These were then compared to the other databases’ results to see how the differences in the values carried through to the assessment.

4.3.3 Questionnaire scores

In **Chapter 3.0: Benchmarks**, the importance of the nature of the benchmark data was emphasised. This section of the analysis aims to explore the data received from the questionnaire for the pilot database.

There are many variables in the questionnaire, with different types of questions. For this section of the analysis the 'summary' variables were focussed on. These all have an ideal score of seven, allowing easier comparison between them. They are also the variables that are used in the BUS summary benchmarks, and have been proven in other questionnaire based methods to be representative of overall performance, although slightly higher scoring. There are 16 in total:

Design, Needs, Space, Image, Meeting Rooms, Storage, Furniture, Temperature Overall in Winter, Temperature Overall in Summer, Air Overall in Winter, Air Overall in Summer, Noise Overall, Lighting Overall, Overall Comfort, Perceived Productivity, Health.

The distribution of the scores for the individual questions is reported in several ways. Initially, this is illustrated in graphical form displaying the distributions visually in histograms, with intervals at a level of accuracy of 0.5.

Descriptive statistics

Statistical analysis is carried out on the data to obtain an idea of how the data relates within itself. This shows the distribution of the scores over the different questions. Descriptive statistics of the distribution of scores are produced through SPSS. There are four main statistics that can be used to classify differences between distributions (Salkind, 2005). They are a measure of central tendency, a measure of variability, a measure of asymmetry, and a measure of flatness. The distribution of the scores is important as it helps to determine how people are scoring the buildings, whether the scores are representative, whether there are outliers, and what statistical tests are appropriate to use. The four specific descriptive statistics that will be presented are the:

- measure of central tendency – mean
- measure of variability – standard deviation
- measure of asymmetry – skewness
- measure of flatness – kurtosis

The mean, as stated earlier, is the average of where the scores lie to give an idea of the typical value of the scores. The standard deviation reflects how the scores are distributed in terms of distance from the mean, a higher standard deviation indicating a more spread out distribution. Through the many examples studied, it was determined that the data from these scales, and questionnaires, can be considered an interval scale, making these statistics the more appropriate to use. As the database is made up of mean scores, there should be minimal outliers, minimising the benefit of using the median in place of the mean.

Skewness relates to the level of symmetry a distribution has. Positive skewness indicates positive skew (scores clustered to the left at the low values), negative values indicates clustering of score at the high end (right hand side of a graph). This is displayed visually below in Figure 24.

Kurtosis deals with the peakedness of a distribution. Positive kurtosis values indicate that the distribution is rather peaked (clustered in the centre), with long thin tails (leptokurtic) and negative kurtosis values indicate a distribution that is relatively flat (platykurtic) (Pallant, 2005), shown below in Figure 25.

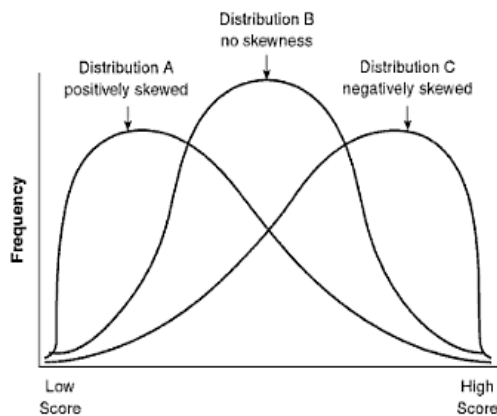


Figure 24: Skewness (Salkind, 2005, p. 97)

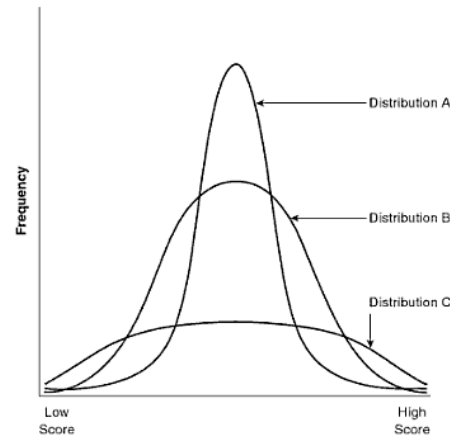


Figure 25: Kurtosis (Salkind, 2005, p. 98)

Normality tests

Skewness and kurtosis are values that can help determine whether the distributions approximate normality. In **Chapter 3.0: Benchmarks** the importance of the normality of the data distribution was considered. For these reasons the distributions of the summary scores are tested, to determine whether the statistics that are used are appropriate, and whether parametric tests will be able to be carried out.

This is done through the use of two methods. Firstly the skewness and kurtosis are considered separately, the methods being very similar for each. A commonly used test to assess the significance of the skewness is to check whether the skewness value calculated is more than two standard errors of skew. The standard error of skew will be calculated through SPSS. The same method is used to consider kurtosis.

The second test of normality is also carried out through the use of SPSS. A Kolmogorov-Smirnov' test was conducted to determine whether the distributions were considered normal.

'The One-Sample Kolmogorov-Smirnov test procedure compares the observed cumulative distribution function for a variable with a specified theoretical distribution, which may be normal, uniform, Poisson, or exponential. The Kolmogorov-Smirnov Z is computed from the largest difference (in absolute value) between the observed and theoretical cumulative distribution functions. This goodness-of-fit test tests whether the observations could reasonably come from the specified distribution' (IBM, 2008).

The test produces a table, with the resultant sig. value being the most important. If the sig. value is above 0.05 the distribution can be considered to approximate normality.

4.3.4 Questionnaire variables

Correlations

Correlations are used to describe the strength and the direction of the relationship between two variables, and provide a numerical summary of this with a correlation coefficient. One that is commonly used is the Pearson correlation coefficient, and this is used in this study. Pearson correlation coefficients can range of -1 to +1. A positive correlation indicates that as one variable increases, so does the other, while a negative correlation indicates that as one variable increases, the other decreases. The strength of the relationship is shown by the size of the absolute value; with 1 or -1 being a perfect correlation, and 0 indicating no relationship.

There are certain assumptions that are required for the use of the Pearson correlation. It is a parametric test so assumes that the dependent variable is measured at the interval or ratio scale, and is normally distributed. Other assumptions are that the observations are independent, not influenced by others, and that the variability in scores for variable X should be similar at all values of variable Y (Pallant, 2005).

This correlation is represented by Pearson correlation coefficients (r values). This figure is comparable to Salkind's rule of thumb for effect sizes, displayed in Table 7.

Table 11: Size effects (Salkind, 2005, p. 129)

Size	Correlation
0.8 - 1.0	Very Strong
0.6 - 0.8	Strong
0.4 - 0.6	Moderate
0.2 - 0.4	Weak
0 - 0.2	Weak or no correlation

In order to get an idea of the representative nature of the buildings, the main summary benchmark methods that BUS use – the three indices, and the BUS rating score: selected variable and BUS rating score: all variable methods– are calculated for this set of buildings using the values from the BUS UK 2006 and NZ 2008 databases, to see how they range over the spectrum to determine the nature of the buildings. The 2008 NZ benchmark was used for this comparison due to availability of the appropriate data.

4.3.5 Building variables

Another variable to consider in defining the proposed benchmarks is to determine whether there should be any splits in the data in terms of the nature of the buildings. This split in the data will be explored through a review of the relevant literature, a study of existing benchmarks, and independent t-tests to determine whether the answers given differ in terms of different variables from the building.

The main focus of this is on those determined in the literature review; mainly the size of the buildings, use, and ventilation methods. There are limitations to this analysis due to the inconsistency between the group sizes and the relatively small sample sizes.

Independent-samples t-test

To determine whether there is a significant difference between the building variables, an independent samples t-test is carried out. This is a procedure for determining whether significant differences exist between two sample means. This procedure is completed through SPSS 16.0 for Windows. The independent-samples t-test identifies whether there is a significant difference between two groups of data.

As with any statistical test, there are certain assumptions that are required for the use of the independent samples t-test. It is a parametric test so assumes that the dependent variable is measured at the interval or ratio scale, and is normally distributed. Other assumptions are that the observations are independent, not influenced by others, and that the variability between the groups is similar (Pallant, 2005).

The results from the analyses produce tables representing the significance of the data. According to 'Levene's Test for Equality of Variances', which is a test for the homogeneity of variance, if the resulting figure is greater than 0.05 it is safe to assume equal variances for all variables (Pallant, 2005).

Independent samples t-tests work on the bases of hypotheses, and the probability that there is a difference between the samples. The limit for when there is a statistical difference is set by the researcher. This set value represents how unlikely a result must be to be considered significant. The traditional level is 0.05 (5%). For the purposes of this analysis, the traditional value is used. Therefore when the 'sig' value in the 't-test' table is less than 0.05 it indicates there is a significant difference.

The effect size is then calculated to determine the degree in which the variables are associated. There are many effect size statistics used but the most common is eta squared. 'Eta squared can range from 0 to 1 and represents the proportion of variance in the dependent variable that is explained by the independent (group) variable' (Pallant, 2005). To calculate the effect size for an independent samples t-test the following formula is used:

Equation 8: Calculating effect size (Pallant, 2005)

$$Eta\ squared = \frac{t^2}{t^2 + (N1 + N2 - 2)}$$

The values for 't' and 'N' are produced by SPSS, and a table can be found in Appendix F. According to Cohen (1988), an eta squared value of 0.01 can be considered a small effect, 0.06 a moderate effect and 0.14 a large effect. These values expressed as a percentage represent the variance in one variable that is explained by another.

4.4 Exploring benchmark format and incorporation into BSRTs

The purpose of this benchmarking analysis is to determine whether the available user questionnaire data can be used to develop a set of benchmarks for the comparison of indoor environmental quality from the occupants' points of view between similar commercial and institutional buildings. The benchmark then needs to be in such a format that it is easily comparable with other building results and can be easily understood.

The objective of this section is;

- To use the data collected to explore methods for determining benchmarks for NZ commercial buildings in terms of user perceptions.

The data from the pilot database is used to explore options of possible benchmarks, and positives and negatives of the options discussed. The methodology of this section involves using the analysis of energy benchmarks, existing questionnaire studies, and current building rating tools and applying some of the methodology from these to the data from the pilot database.

Specific data from a high performing and a low performing building from this database were chosen to test the sensitivity of the options.

The benchmarks are explored through two formats; the first being a summary of the building through the nature of the scores themselves, referred to as Absolute benchmarks. The second relates these scores to the rest of the database, referred to as Relative or Combined benchmarks.

Once the different benchmarks are explored, investigations are made as to how these could potentially be incorporated into building rating tools, such as those mentioned previously in **Section 3.5: Current building rating tools**.

5.0 ANALYSIS OF SCORES

This study explores the nature of the buildings and the questionnaire results used to form the pilot database. Firstly, this section looks at the design of the buildings. Classification of the design allows a view of how these buildings relate to the overall NZ building stock, and the differences between them. Following this is a comparison between the different databases used at present and how the buildings rate according to these summary methods. An exploration of the type of data collected, and the relationship between the results collected and the different building types is also described, and summary paragraphs are provided at the end of each section.

The number of respondents per building ranged from 11 to 439. This is a huge difference in sample size, so the mean scores for the smaller buildings will be more sensitive to extreme scores than those in the larger buildings with a greater number of occupants. The mean number of respondents over the buildings is 105, with a median of 74.

5.1 Classification of design

The pilot database consists of 30 buildings from around New Zealand. Their designs were categorised using the method described in **Section 4.3.1: Classification of design**. The buildings are very different in their size, and design, but are all commercial and institutional in use and are situated in New Zealand. Classifying the buildings in this way allows some comparison between them, and indicates how representative they are of the New Zealand building stock.

5.1.1 New Zealand building stock

The proposed benchmarks will be New Zealand specific and ideally should be cognisant of the makeup and distribution of the building stock of New Zealand.

The following graphs show the distribution of the building types in New Zealand. The New Zealand building stock is relatively small and is comprised largely of residential buildings. The two main commercial cities in New Zealand are Auckland and Wellington; there are more buildings in the North Island than the more sparsely populated South Island. Auckland has by far the most commercial buildings, in floor area and number, as would be expected.

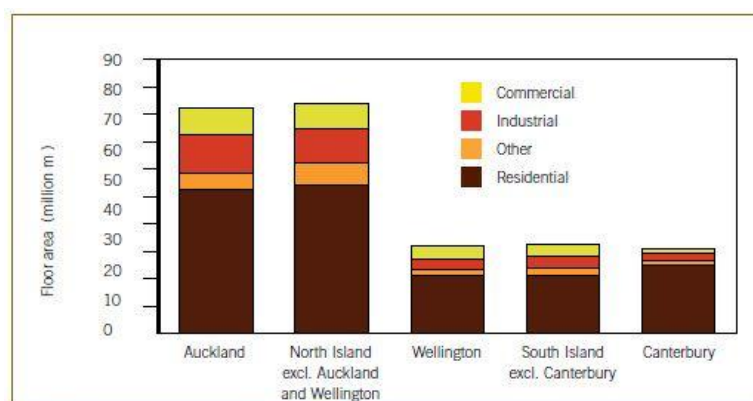


Figure 26: Distribution of building types by floor area (Isaacs, 2008, p. 45)

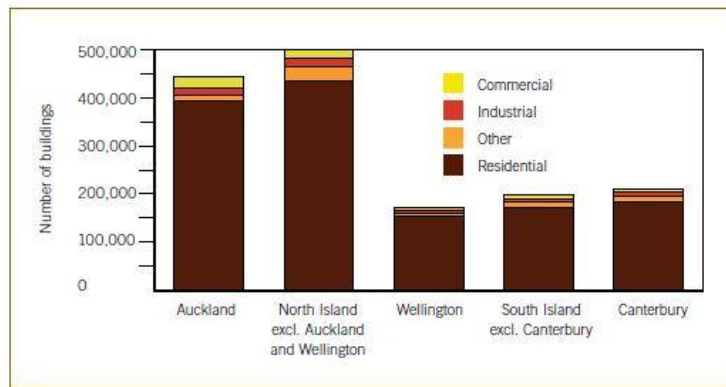


Figure 27: Distribution of building types by numbers (Isaacs, 2008, p. 45)

5.1.2 Pilot database

Out of the 30 buildings in the pilot database there are 23 commercial buildings and 7 institutional. The commercial buildings are mainly made up of office space with meeting rooms, and are predominantly utilised by the same people doing similar tasks. The institutional buildings also include lecture rooms and are used by both students and academic staff. They are both primarily places where sedentary office work is undertaken, and where people spend long periods of time. There are an obvious larger number of commercial buildings than institutional buildings, but it was deemed that these could be looked at together due to the similar main use in both building types, and the fact that the questionnaires used in the analysis were completed by permanent staff. The reasoning behind this was based on BUS where they use groups of 'non-domestic buildings of all types with permanent occupants' (Building Use Studies, 2009, p. 1).

The buildings in the database are situated around New Zealand, but predominantly in Wellington and Auckland. The following graph shows the distribution. The graph shows that just over half of the 30 buildings, 53%, are situated in Wellington, followed by 33% in Auckland. These are the two main commercial cities in New Zealand so it is appropriate that these are the largest proportions. Comparing the above graphs these are the main centres that were separated from the more general 'North Island' grouping. There is an inconsistency however in the proportions between the two compared to the general distribution of buildings throughout the country, shown in Figures 26 and 27 above. The main reason for the larger number of buildings in Wellington is that the main two sources of building surveys, Victoria University of Wellington, and E Cubed Building Workshop, are based in Wellington and therefore have greater access to the buildings in this area. The remaining four buildings are situated around the country and include the only three buildings from the South Island.

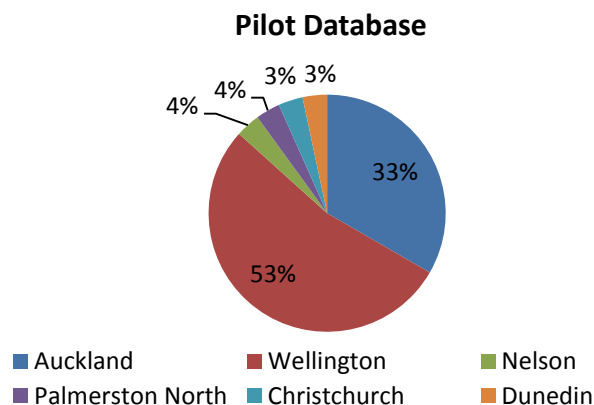


Figure 28: Proportions of pilot buildings around NZ

In terms of different floor areas, the buildings were split into the same five floor-area categories as BEES. The categories are determined by splitting the buildings evenly by floor area. The following graph shows how this group of buildings was distributed over them.

In terms of proportions of the NZ building stock, 'by count, 87% of non-residential buildings are under 1,500m² in floor area, but the remaining 13% by count represent 60% of the floor area' (Isaacs, et al., 2009, p. 2).

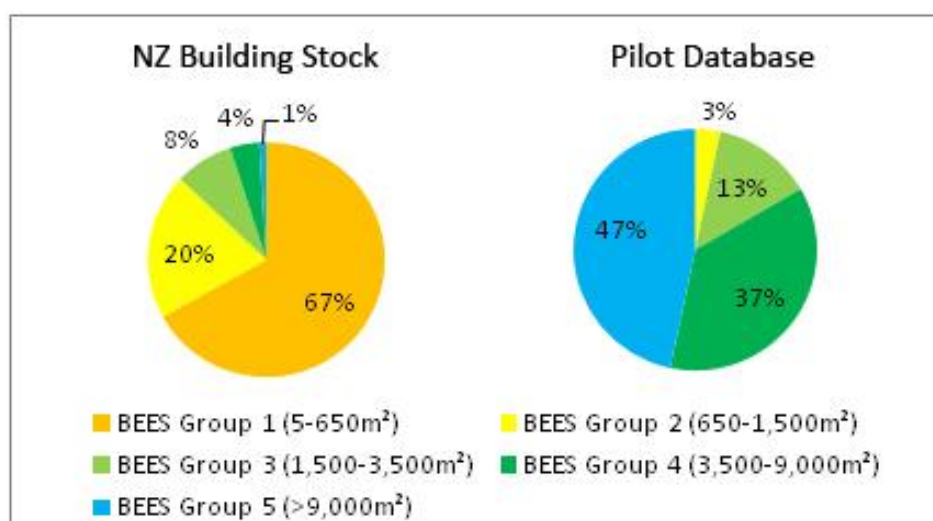


Figure 29: Comparison between NZ building stock and pilot database is terms of size

The vast majority of the pilot database buildings are of the larger size, in BEES categories 4 and 5. Comparison between the NZ building stock and the pilot database shows that in what was deemed the largest grouping, in terms of building numbers, of non-residential buildings in NZ this sample has none, with the smallest buildings that have been surveyed being from 650m² and above (BEES Group 2). The smallest group in the NZ building stock is the largest of groups here from this sample. This indicates that in terms of building size this sample is not representative of the NZ building stock, and that the 30 buildings surveyed are mainly in the minority. The larger buildings generally have greater occupancy numbers, and are therefore more attractive to survey as they mean a larger sample of responses for a particular building. Also, owners of larger buildings have a bigger investment and hence interest in the performance of buildings and are likely to have resources to follow this through.

The characteristics of the buildings which have been identified as relevant are those which previous studies have determined to be significant or important in the description of the buildings.

TOBUS, one of the questionnaire based methods that were studied in **Section 2.3: Examples of questionnaires used** identifies the differences in design between their building sample. They present the results in a table, shown below.

Table 12: Example of building classification: TOBUS methodology (Caccavelli & Gugerli, 2002, p. 6)

Buildings surveyed during the TOBUS method and software development

Country	Age of building			Size of building (m ²)			Building height (storeys)			Office type		Category (prestige)			Mixed use	
	<1930	1930–1960	>1960	<1000	1–2000	>2000	<3	3–6	>6	Open	Cell	Low	Normal	High	Yes	No
CH1			X			X			X		X		X			X
CH2			X			X			X		X		X			X
CH3	X					X		X			X		X			X
DK1			X		X		X				X		X			X
DK2			X			X		X			X		X			X
DK3	X					X		X			X			X		X
FR1			X			X			X		X		X			X
FR2		X				X		X			X	X			X	
FR3			X	X			X				X		X			X
GR1			X		X				X		X			X	X	
GR2			X		X			X			X		X		X	
GR3	X			X			X				X		X			X
NL1			X			X		X			X		X		X	
NL2			X			X		X			X		X			X
NL3			X		X			X			X		X			X

Adopting this methodology, but shifting the bands makes it possible to see how the buildings differ from each other specific to this sample. As well as the building size that was considered earlier, the buildings can be classified in terms of building height, to give an idea of the overall shape and size of the buildings. The building age will have an effect on the type of design, as trends in the building industry change over time. For example in recent years there has been much more focus on energy efficient design, where previously there was a large trend for fully sealed, air conditioned buildings. Almost half of the buildings surveyed have six or more floors, and the majority of them were built or significantly refurbished since 2000.

Table 13: Classification of design features

	Building Use	Building Age			Building Height (floors)			BEES Floor area group	Ventilation	Design Intent
		< 1970	1970-2000	>2000	<3	3-6 floors	>6			
Building 1	Commercial			X			X	4	MM	Sustainable
Building 2	Commercial			X			X	4	AC	Sustainable
Building 3	Institutional			X	X			2	NV	Sustainable
Building 4	Commercial			X		X		4	MM	Sustainable
Building 5	Institutional			X		X		5	ANV	Sustainable
Building 6	Institutional		X			X		4	MM	Sustainable
Building 7	Institutional		X		X			3	NV	Sustainable
Building 8	Institutional			X	X			3	AC	Sustainable
Building 9	Institutional		X				X	5	NV	Sustainable
Building 10	Commercial		X				X	5	MM	Conventional
Building 11	Institutional			X		X		4	MM	Conventional
Building 12	Commercial		X			X		5	MM	Conventional
Building 13	Commercial			X		X		5	AC	Conventional
Building 14	Commercial	X				X		4	NV	Conventional
Building 15	Commercial		X				X	5	AC	Conventional
Building 16	Commercial		X				X	5	AC	Conventional
Building 17	Commercial		X				X	5	AC	Conventional
Building 18	Commercial		X				X	5	AC	Conventional
Building 19	Commercial		X				X	4	AC	Conventional
Building 20	Commercial		X				X	4	AC	Conventional
Building 21	Commercial		X			X		4	AC	Conventional
Building 22	Commercial		X				X	5	AC	Conventional
Building 23	Commercial		X				X	5	AC	Conventional
Building 24	Commercial		X			X		3	MM	Conventional
Building 25	Commercial			X		X		4	AC	Sustainable
Building 26	Commercial	X					X	5	AC	Conventional
Building 27	Commercial	X					X	5	AC	Conventional
Building 28	Commercial			X		X		5	AC	Sustainable
Building 29	Commercial			X			X	4	MM	Sustainable
Building 30	Commercial			X		X		3	AC	Sustainable

There have also been studies carried out focussed on sustainable buildings, either looking at the sustainable buildings by themselves or a comparison between them and more conventional buildings (Leaman & Bordass, 2007; Leaman, Thomas, & Vandenberg, 2007; Baird, 2010; Baird, Leaman, & Thompson, in press). One example of this is HOPE where half of the buildings selected had to include energy saving measures (Bluyssen, Aries, & van Dommelen, 2011).

In this pilot database of 30 NZ buildings there are a mixture of ‘sustainably designed’ and other more conventionally designed buildings. Almost half of the buildings have some sort of sustainability credentials. All of the buildings were designed with the explicit intent of being environmentally friendly and low energy users, as described in **Section 4.3.1: Classification of design**. Table 14 displays the buildings classed as having a

sustainable design intent and show an example of some sustainable design features they have incorporated, and awards the buildings have won.

Table 14: Sustainable buildings' design features and awards

	Open Plan	Atrium	Exposed Thermal Mass	Shading devices	High tech glazing	Efficient lighting	BMS	Rain water collection	Awards
Building 1	X	X			X	X	X	X	Resene NZ Awards, Wellington Civic Trust Award, Property Council of NZ Commercial Office Energy Efficiency Award. Ministry for the Environment ESD Award
Building 2	X					X	X		Innovate NZ Award - Merit, Genesis Energy Commercial Award - Commended
Building 3	X		X	X		X	X		Energy-Wise Commercial Buildings Award
Building 4		X	X	X	X			X	Energy-Wise Commercial Buildings Award
Building 5	X	X	X	X	X				Property Council of NZ Award, an NZIA National Award, an ACENZ Silver Award and an EECA Energy Efficiency Award.
Building 6	X	X	X	X			X		
Building 7	X	X	X	X					
Building 8	X		X	X	X				NZIA Supreme Award, Innovate NZ Award - Silver
Building 9		X	X	X			X		Gold Award of the Association of Consulting Engineers of NZ
Building 25	X	X	X	X	X	X	X	X	NZGBC 5-stars, Concrete3 Sustainability Award of Excellence, NZ Engineering Excellence Award for Sustainability, NZIA NZ Architecture Medal Finalist, ACENZ Gold, Property Council Awards, SBN Local and National and Innovation Award
Building 28	X	X	X	X	X	X	X	X	Innovate NZ Award - Merit, Two NZ Property Council Awards, Two NZIA Resene Awards
Building 29	X		X	X	X	X	X		NZGBC 5-stars, NZIA Sustainable Architecture Award
Building 30	X		X	X				X	NZGBC 4 stars

A common feature of many of these studies is the ventilation method. The buildings were split into three ventilation methods; Natural Ventilation (NV), Mixed Mode (MM) and Air Conditioned (AC), as shown in Table 13.

The **Natural Ventilation (NV)** category is made up of two types; full natural ventilation and advanced natural ventilation. The database includes five such buildings.

Mixed-Mode (MM) buildings consist of three types of systems; zoned, concurrent and changeover. There are eight buildings in this category.

Air Conditioned (AC) buildings consist of fully air-conditioned buildings and buildings that are completely sealed. The windows in these buildings are not intended to be opened. There are 17 buildings in this category.

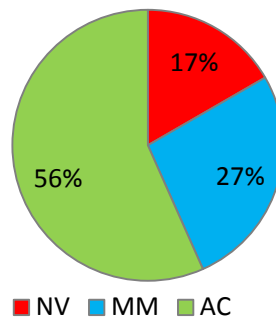


Figure 30: Ventilation distribution

5.2 Comparison of benchmarks and databases

This section focuses on the BUS benchmarks that have been used to assess the buildings previously, in comparison to the pilot database. This first contains a direct comparison between the mean scores and the confidence intervals of the summary variables. This is followed by a more inclusive method using all questions in the questionnaire and their distance to ideal scores. The mean scores of the pilot database are then compared to the results that were in the original reports in regards to the 'Better', 'Similar' and 'Worse' scores and then in comparison to the separate benchmarks. The comparison of the buildings in terms of the BUS benchmarks reveals how they are being benchmarked and presented currently and how the databases differ.

5.2.1 How the pilot database's averages compare to the current BUS benchmark being used

Looking at all of the summary questions that have an ideal score of seven, a comparison can be made between the mean BUS UK 2006 and NZ 2010 benchmarks compared to the mean of the pilot database. For this the most recent of all of the available data was used. The error bars displayed are the 95% confidence levels; giving 95% surety that the mean from the population will fit between the limits from these samples.

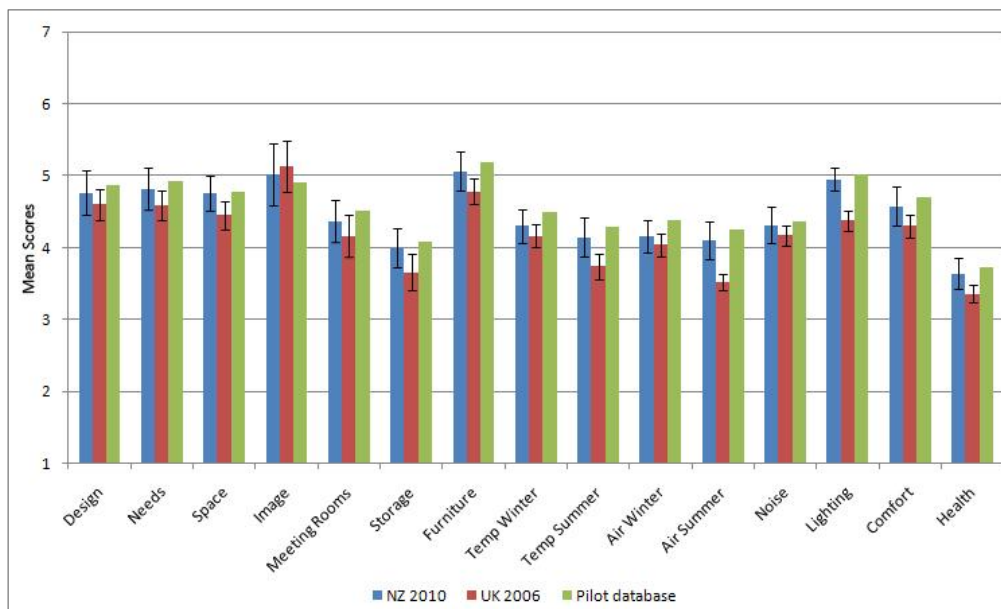


Figure 31: Comparison of databases according to summary variables

All three of the building groups are performing above the midpoint, of four, for the majority of the variables. They generally follow the same trends as each other, with the UK data set rating the lowest the most often. The pilot database rates the highest the most often, but is rating very close to the BUS NZ 2010 benchmark. The sole exception to this pattern is *Image*, where the BUS UK 2006 benchmark is higher than both the BUS NZ 2010 benchmark, and the mean from the pilot database. There is the difference in sample size that needs to be kept in mind, as the NZ benchmarks are quite a lot smaller than the UK. The BUS NZ 2010 benchmark is based on 24 buildings, the BUS UK 2006 benchmark is based on 50 and the pilot database is made up of 30 buildings (24 for *Air Overall in Winter* and *Air Overall in Summer*).

Temperature Overall in Summer, *Air Overall in Summer*, *Health*, and *Storage* all feature mean scores below the midpoint indicating that these may be low performing attributes of the buildings.

Perceived Productivity, although a summary variable, is scored on a different 9-point scale ranging from -40% to +40%. In terms of the rating between benchmarks the order is much the same with the pilot database scoring higher than both the UK 2006 benchmark and the NZ 2010 benchmark. There is no significant difference between the UK 2006 benchmark and the NZ 2010 benchmark, as the UK confidence interval is completely within the confidence limits of the NZ 2010 benchmark.

The UK confidence interval is smaller in every case – this could be due to the larger sample size, as it is more representative of the population. In many of the cases the UK confidence intervals fit completely within the NZ confidence limits. The largest range of confidence intervals is for *Image* in both the BUS NZ 2010 benchmark, and the BUS UK 2006 benchmark

The databases are small in comparison to the overall population sizes, but there are questions where according to the confidence intervals the variables are scoring differently, which reflects the need for country-specific benchmarks. For *Air Overall in Summer* and *Lighting Overall* the confidence intervals are completely outside each other's bands, which indicate that the population mean for these would be different. These are the only variables where this is the case.

Comparison of the databases in relation to all questions

In order to include all of the 45 questions in the questionnaire, regardless of the type of scale they employed, the mean scores were compared to the ideal score of the corresponding question. Looking at this in another way and including all of the sub questions can also give an overall indication of the relationship between the three datasets. In this case all of the questions were used and they were ranked from 1 (closest to ideal) to 3 (furthest from ideal). This is using only the mean, so the differences may be small and non-significant in terms of the confidence intervals.

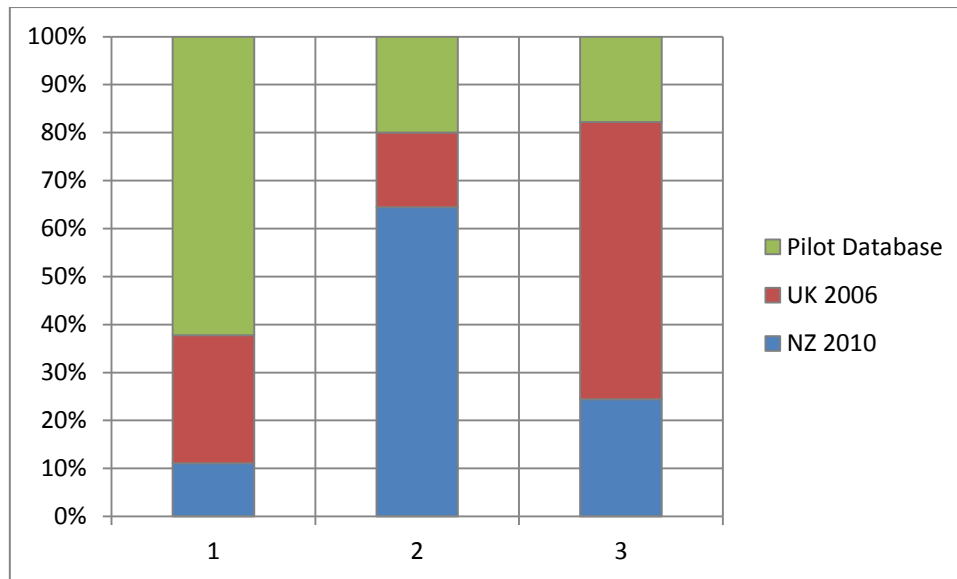


Figure 32: Comparison of databases including all questions

This shows that for 62% of the questions the pilot database is scoring the closest to ideal. This is consistent with what was shown in the previous graph with the summary variables; of the pilot database scoring higher, and the BUS UK database scoring lower. It also indicates that this trend follows on for the majority of the questions and is not isolated to the summary variables.

5.2.2 Comparison of buildings in terms of current BUS benchmark and dataset

As described in **Section 3.4: Assessment of BUS benchmarks** the buildings were individually compared to the Building Use Studies (BUS) moving benchmark, either NZ specific or the UK dataset depending on when they were surveyed. They were rated as being *'Better'* or *'Worse'* than the benchmark when the specific mean score for the building was outside the 95% confidence levels of the BUS buildings and midpoint, and *'Similar'* when within the range.

This comparison indicates whether the buildings are more 'occupant friendly' spaces to be in, compared with an average. As the BUS benchmark is dynamic the range of critical limits of the benchmark change every year when it is updated.

Figure 33 illustrates the overall performance of the 30 buildings, using 16 summary variables as an indication, of the comparison of the individual building reports that were completed over time. This is the result for each variable that the building owner would have received in the report, to give an idea of the distribution of results that were received. The percentages on the y-axis represent the number of buildings that rate *'Better'*, *'Similar'* or *'Worse'* than the benchmark, with the summary variables along the bottom. These ranged over different datasets, NZ and UK, and were from 2001 to 2010. Those surveyed in 2007 or later were compared to the NZ database and the earlier period were compared to the UK database.

The difference in benchmarks used over this period of time is due to the dynamic nature of the benchmarks and the attempt to get the most relevant comparison at the time of the survey. It is only due to the fact that these buildings are being used to form another benchmark that the individual building scores are being reviewed and compared to a static benchmark.

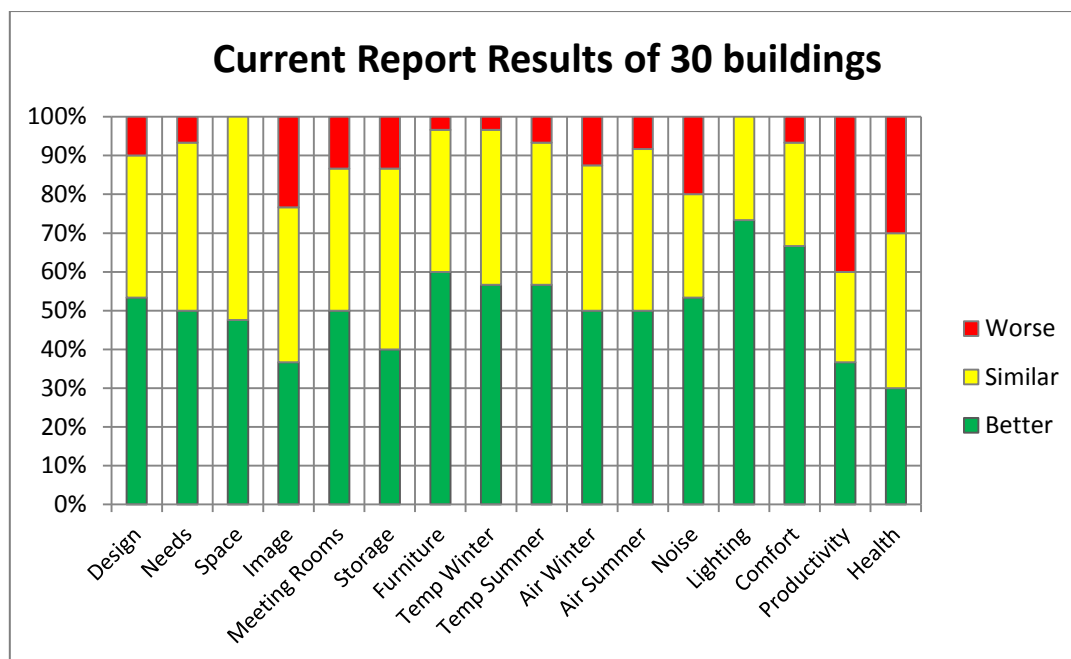


Figure 33: Summary variable comparison to benchmarks regardless of time of study

Figure 33 illustrates a range of results over the different variables. The variable with the highest number of buildings scoring better than the benchmark is *Lighting Overall*; 73% were 'Better', 27% were 'Similar' and none were 'Worse' in this case. The variable with the least amount of buildings scoring 'Better' than the benchmark is *Health*, with 30%, while *Perceived Productivity* had the largest number of buildings reported as being 'Worse' than the benchmark, with 40%. In terms of the indoor environment variables (*Overall Comfort*, *Lighting Overall*, *Noise Overall*, *Temperature Overall* and *Air Overall* in Winter and Summer), all of the variables had at least 50% of the buildings scoring 'Better' than the benchmark, so overall they are not doing badly. Out of these the lowest number was for *Noise Overall* with 53%. *Noise Overall* also featured the highest number of buildings scoring 'Worse', with 20% of the buildings.

In order to get an idea of the individual performance of the buildings in terms of a more static benchmark, the individual building scores were compared to the same benchmark with the same critical limits. This was done separately for the BUS UK 2006 benchmark and the BUS NZ 2010 benchmark. In the first part of this chapter the means of the different benchmarks were compared and showed that they were relatively close. Comparing the individual buildings' scores to both of the BUS benchmarks shows whether the slight differences in them have an effect in terms of the reporting.

As the majority of the buildings in this database were originally compared to a UK benchmark, this was considered first. This database is made up of the last 50 UK buildings that have been assessed using this methodology. The buildings in the pilot database that were compared to this benchmark originally were surveyed in the years from 2001 to 2006. The most recent of these, the 2006 benchmark, was used as a comparison for all of the buildings. The comparison of this to the individual building scores is shown below, in Figure 34.

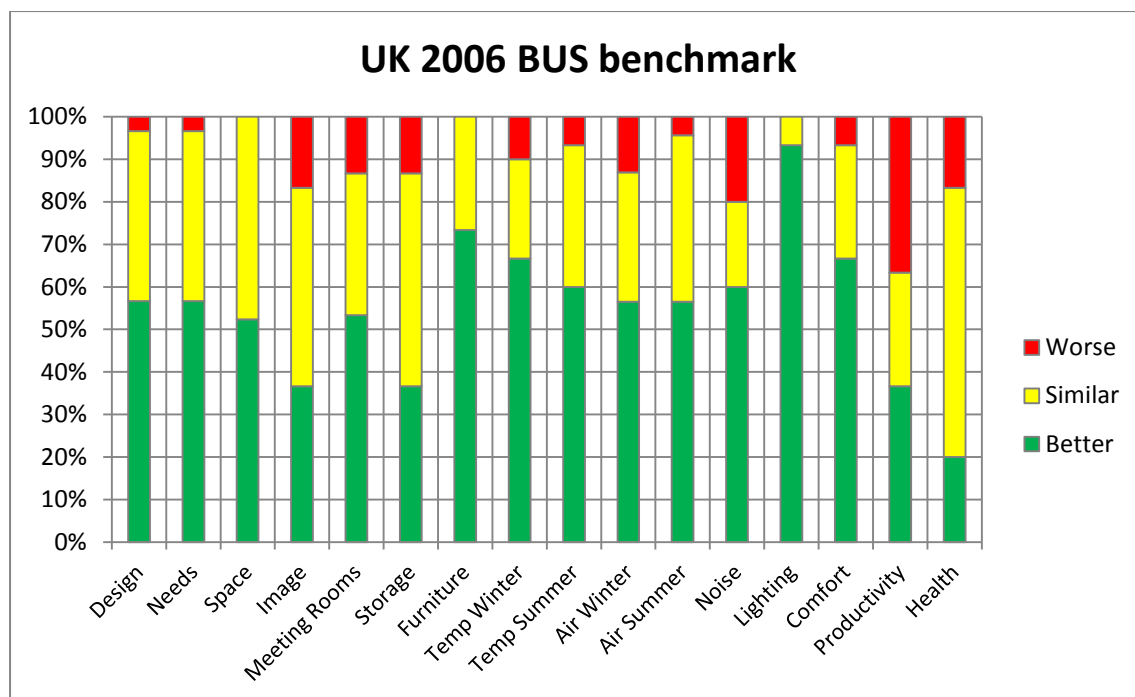


Figure 34: Summary variable comparison to UK 2006 benchmark

The result has some similar and some different aspects to what was reported and presented in the previous graph. The pattern stays relatively consistent over all of the variables with generally only the proportion changing slightly.

The largest source of dissatisfaction remains with the *Perceived Productivity* of the building, with 37% of buildings rating 'Worse', and the smallest number of buildings scoring 'Better' than the benchmark stays consistent in terms of *Health*, with 20%, although this has decreased by 10%.

Lighting Overall is still the variable with the highest number of buildings scoring 'Better' than the benchmark, with 93%, and this has increased with all but two buildings scoring 'Better'.

The most recent surveys carried out using the BUS NZ benchmark were the buildings surveyed during this study, Aorangi House and Waitakere Central One, in 2010; therefore the critical limits from the BUS NZ 2010 benchmark were used as basis for determining the 'Better', 'Similar', 'Worse' ratings. This benchmark database consisted of the average scores of 24 buildings. This is due to the limited number of surveys that have currently been undertaken in New Zealand. It should be kept in mind that potential differences between these results and the UK comparison above is not only the country the benchmarks are from but also the much smaller sample size that could be affecting the results.

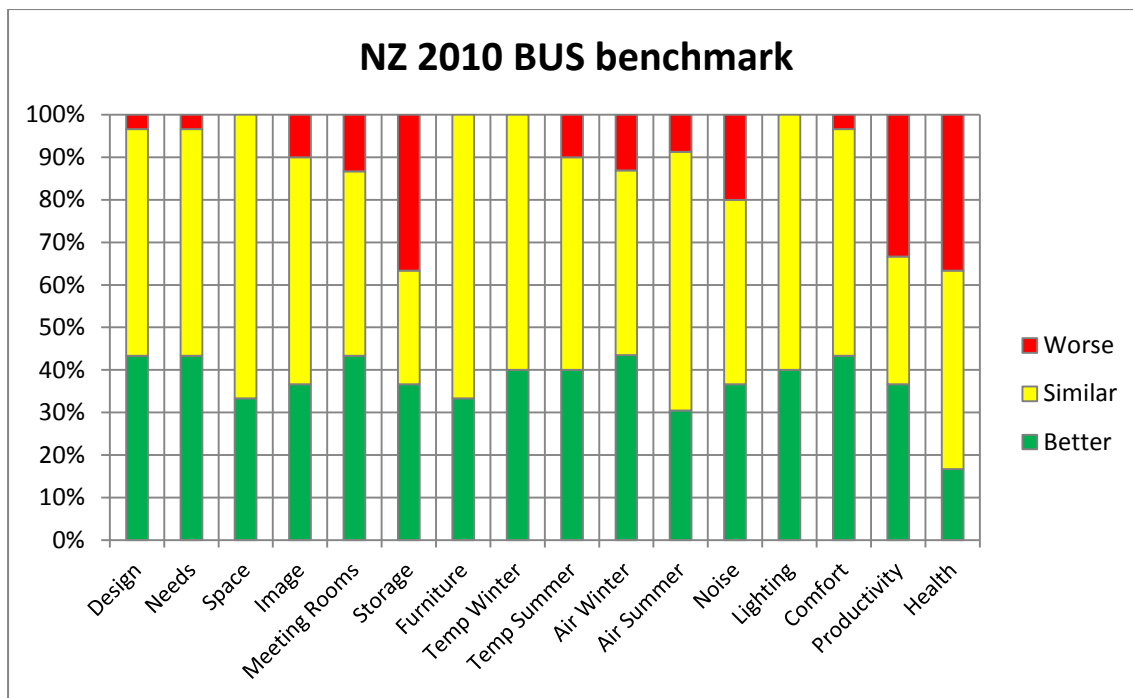


Figure 35: Summary variable comparison to NZ 2010 benchmark

There is quite a large difference compared to the UK benchmark comparison. Generally, there are fewer buildings performing 'Better', and more performing similarly by comparison. The number of buildings performing 'Better' than the benchmark for *Health* has reduced even more, emphasising that the occupants perceive the large majority of these buildings to be similar to or less healthy than the average building. This difference in distribution indicates that there is a difference between the datasets from different countries as they each have their own unique issues. The fact that this group of buildings are situated in NZ means that the NZ benchmark is more applicable. The other reason for the similarity is that some of the building scores could be part of both this pilot database and the BUS NZ 2010 datasets to make up the benchmark.

5.2.3 Summary

The BUS benchmark buildings and the pilot database score relatively alike, with the pilot database performing slightly better than the other two. Even though the differences are slight, when carried through to an assessment on individual buildings, they are enough to make a difference to the ratings.

The ratings show that there is a common trend throughout the benchmarks in terms of the variables that perform highly and those that have room for improvement, and it is reassuring to see that the pilot database follows a similar trend. The two variables that had significant differences between the databases were *Air Overall in Summer* and *Lighting Overall*, with the UK database scoring lower. This could be due to building size, or to different climatic conditions (e.g. short days in UK winters, or clearer air in NZ). There is a large difference in the 'Better', 'Similar', 'Worse' ratings for these variables using the different databases, but it is no more significant than the other variables.

5.3 Overview of scores of the buildings

This section is made up of four subsections. The first assesses visually the score distributions of different summary variables to determine patterns and to obtain an understanding of how people are rating these buildings. The next two sections look in more detail at the skewness and kurtosis characteristics of the distributions to see whether these are significant, and finally a general normal distribution test to back up the skewness and kurtosis findings and to determine which statistical tests are most appropriate to be carried out on the data is described.

5.3.1 Distribution

Distributions of the summary variables are presented in a consistent way so that initial assessments can be made visually. This consists of an example of how the question is phrased in the questionnaire, with the ideal score highlighted in green. This is followed by the frequency distribution graph. The main features of these are located and labelled in the following image (Figure 36). This is presented for the 16 summary variables determined in **Section 4.3.3: Questionnaire scores**.

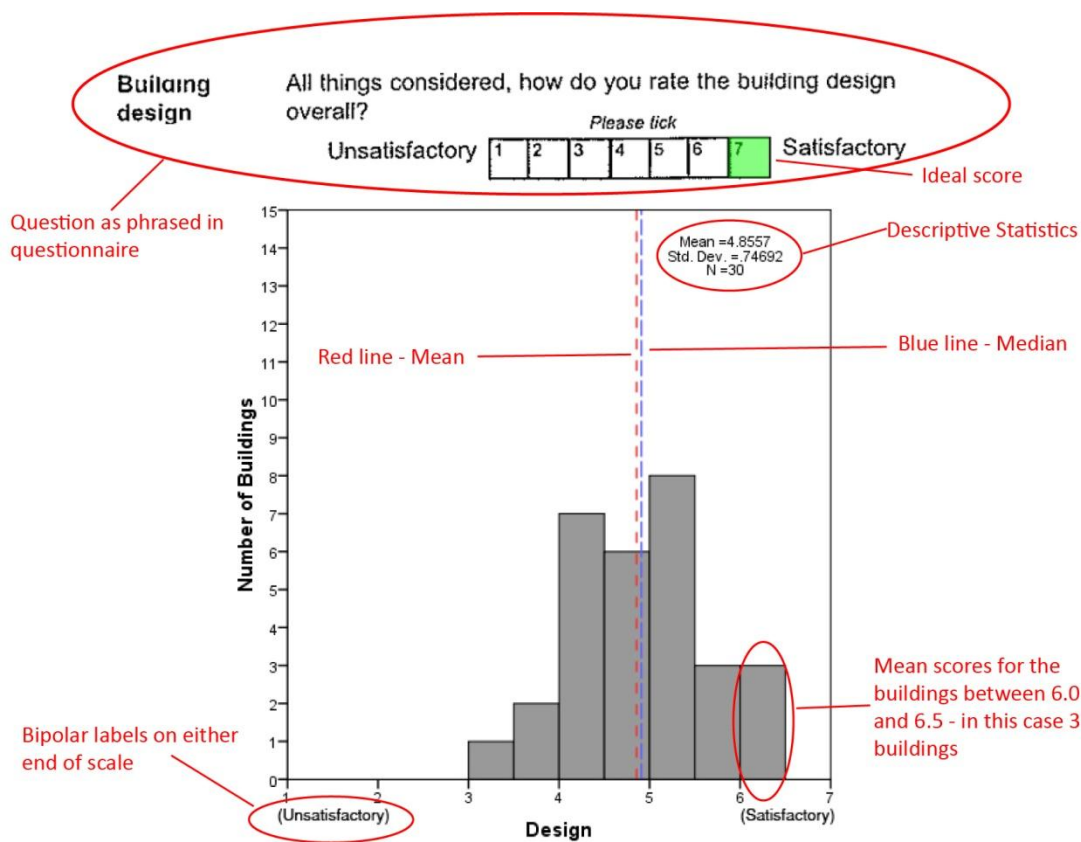


Figure 36: Display example

Design

This question in the POE questionnaire was structured as below, with the ideal answer highlighted in green. The intention of this question is for the occupants to state how they find the design of the building overall.

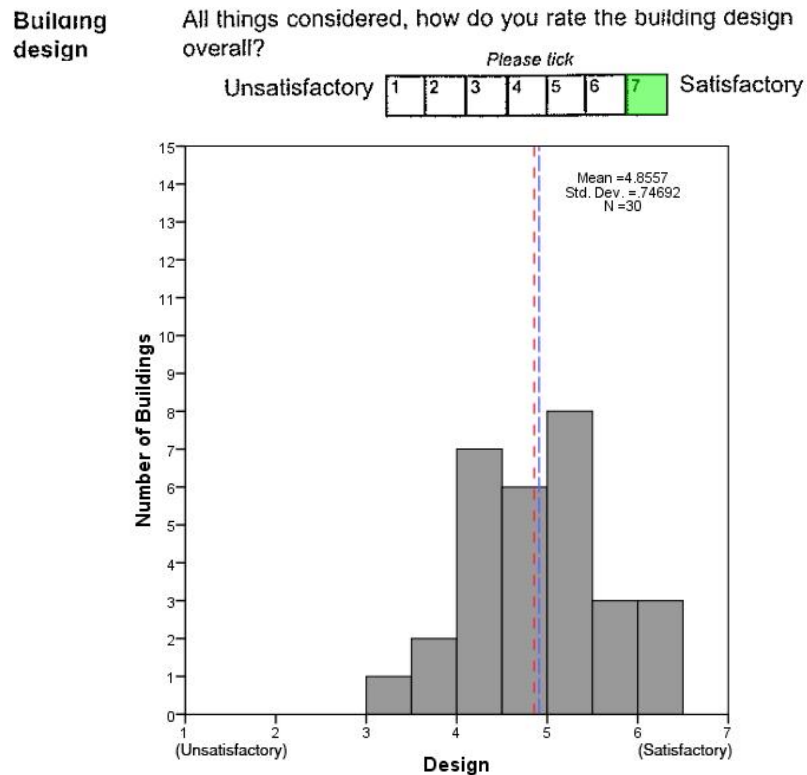


Figure 37: Frequency distribution of mean Design scores

Table 15: Descriptive statistics of Design scores

Design				
Sample Size (N)	Mean	SD	Skewness	Kurtosis
30	4.856	0.747	-0.091	-0.735

Figure 37 displays the frequency distribution of the mean scores for each building. Both the mean and the median are above the midpoint of the scale; so indicates that for the majority of the buildings the occupants judge the design to be satisfactory. Results for all but three buildings are above the midpoint of the scale. The distribution has a negative skewness as the distribution has a slight leaning towards the higher values. The distribution is relatively flat, with no one specific peak, resulting in the negative kurtosis.

Needs

The question on needs from the POE questionnaire is structured in the following way. This is asking the occupants how well the building as a whole works for the tasks they need to complete while in it.

Needs In the building as a whole, do the facilities meet your needs?

Please tick

Unsatisfactory

1	2	3	4	5	6	7
---	---	---	---	---	---	---

 Satisfactory

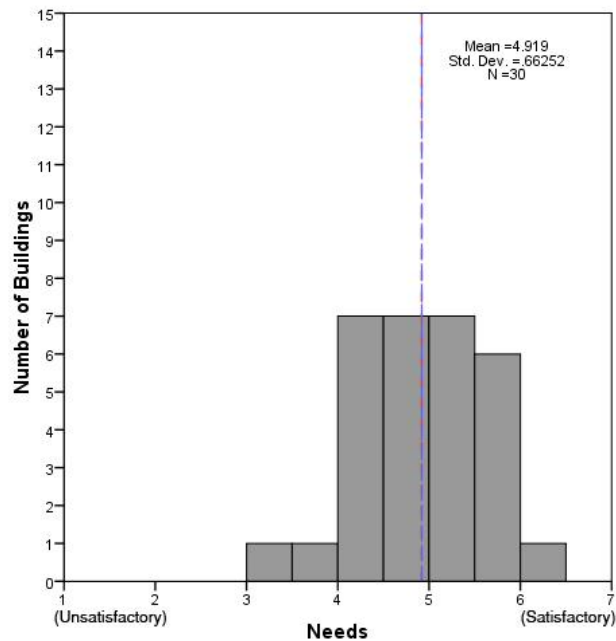


Figure 38: Frequency distribution of mean Needs scores

Table 16: Descriptive statistics of Needs scores

Needs				
Sample Size (N)	Mean	SD	Skewness	Kurtosis
30	4.919	0.663	-0.326	-0.343

This has similar properties to *Design*. Figure 37 displays visually the kurtosis and skewness of the data. The majority of the scores (all but two) are above the midpoint of the scale, resulting in the higher mean value of 4.92. The occupants rate all but two buildings to be satisfactory at meeting their needs. The distribution is still towards the flatter end of the spectrum, but not to the same extent as *Design*.

Space

The occupants are asked to rate how the space in the building is used overall; there are follow up questions that deal with space at their individual work stations.

Space

In the building as a whole, do you think that space is used ...?

Please tick

Ineffectively overall	1	2	3	4	5	6	7	Effectively overall
--------------------------	---	---	---	---	---	---	---	------------------------

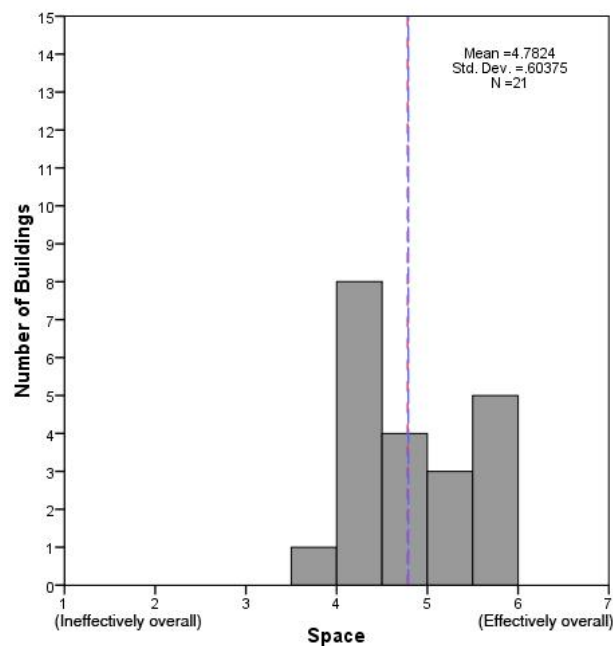


Figure 39: Frequency distribution of mean Space scores

Table 17: Descriptive statistics of Space scores

Space				
Sample Size (N)	Mean	SD	Skewness	Kurtosis
21	4.782	0.604	0.277	-1.282

Space in the building has a different distribution from some of the others. Due to availability of data there were only results from 21 of the 30 buildings that make up the pilot database. The mean of these is relatively high, with all but one building scoring above the midpoint of the scale and receiving a positive rating from the occupants. The kurtosis is relatively high, towards the flatter distribution.

Image

The intention of this question is for the occupants to make an assessment on how they think the building presents itself to visitors.

Image How do you rate the image that the building as a whole presents to visitors...? *Please tick*

Poor ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☒ 7 Good

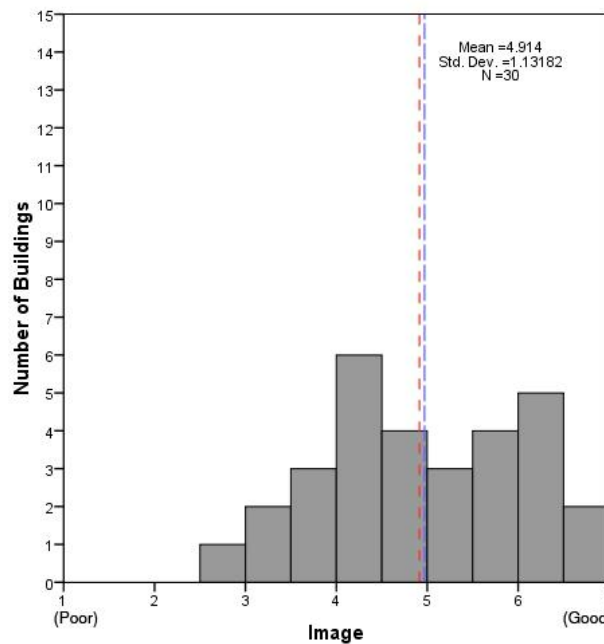


Figure 40: Frequency distribution of mean *Image* scores

Table 18: Descriptive statistics of *Image* scores

Image				
Sample Size (N)	Mean	SD	Skewness	Kurtosis
30	4.914	1.132	0.002	-1.063

Image is a particularly subjective question. The distribution has the largest variability. There is no strong trend towards one answer and the scores are distributed relatively evenly over the range, resulting in the skewness value being very close to zero. This is also shown through the high negative kurtosis, and variance. The mean score of 4.914 being near the 5-point of the scale indicates that for most of the buildings the occupants rate the building as being satisfactory; 24 buildings rating above the midpoint of four.

Meeting Rooms

This question asked the occupants to rate the availability and quality of meeting rooms. There is also a space in this question for comments about them.

Availability of meeting rooms

Please tick

Unsatisfactory 1 2 3 4 5 6 7 Satisfactory

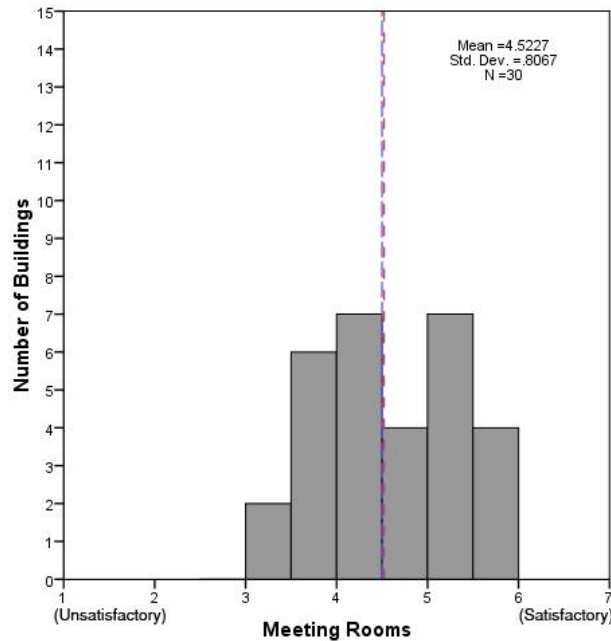


Figure 41: Frequency distribution of mean Meeting Rooms scores

Table 19: Descriptive statistics of Meeting Rooms scores

Meeting Rooms				
Sample Size (N)	Mean	SD	Skewness	Kurtosis
30	4.523	0.807	-0.005	-0.944

The skewness being very close to zero means that there is almost an even number of buildings on either side of the mean. The mean itself is above the midpoint of the scale, and the kurtosis shows a relatively flat distribution. This could largely be due to there being two main peaks in the distribution, but the sample may have to be larger in order for this to be fully realised. These peaks are at just above the midpoint and then at just above the 5 point of the scale.

Storage

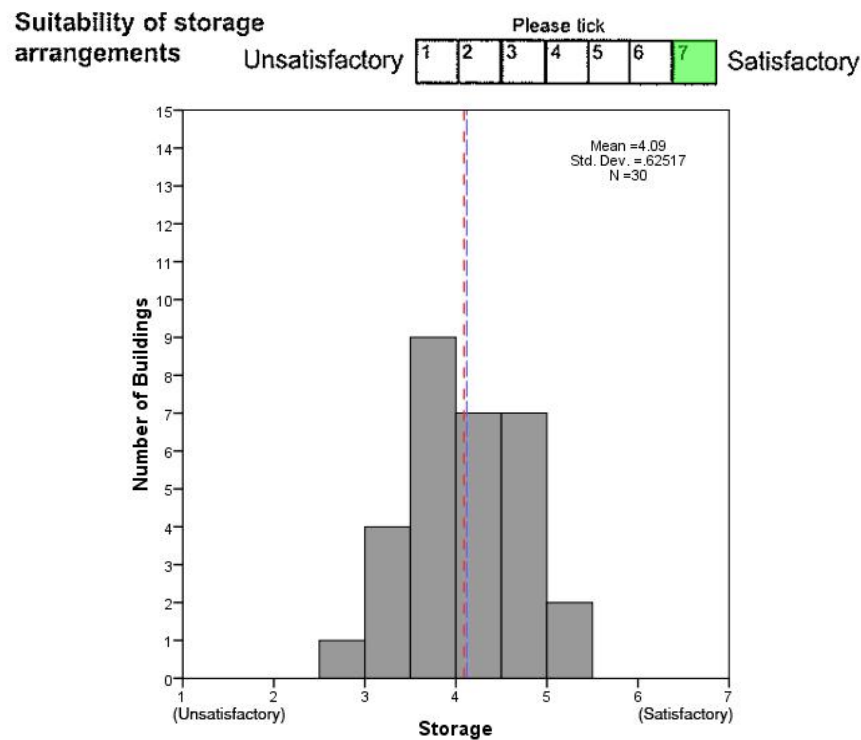


Figure 42: Frequency distribution of mean Storage scores

Table 20: Descriptive statistics of Storage scores

Storage				
Sample Size (N)	Mean	SD	Skewness	Kurtosis
30	4.090	0.625	0.063	-0.720

The mean for this question and this group of buildings is very close to the midpoint of the scale. The skewness score is positive as there are more buildings towards the lower end of the distribution. The kurtosis is negative as the scores were generally quite evenly distributed over the range. This is so far the lowest scoring variable in the pilot database.

In previous studies this storage variable has been shown to be one of the problematic areas of buildings' designs.

'As evidenced by the nature of the comments and to some extent the scores for these factors, Noise and Storage issues were by far the commonest source of complaint' (Baird, 2010, p. 20).

Furniture

Furniture How do you rate the usability of the furniture provided at your desk or normal work area ...?

Please tick

Very poor 1 2 3 4 5 6 7 Very good

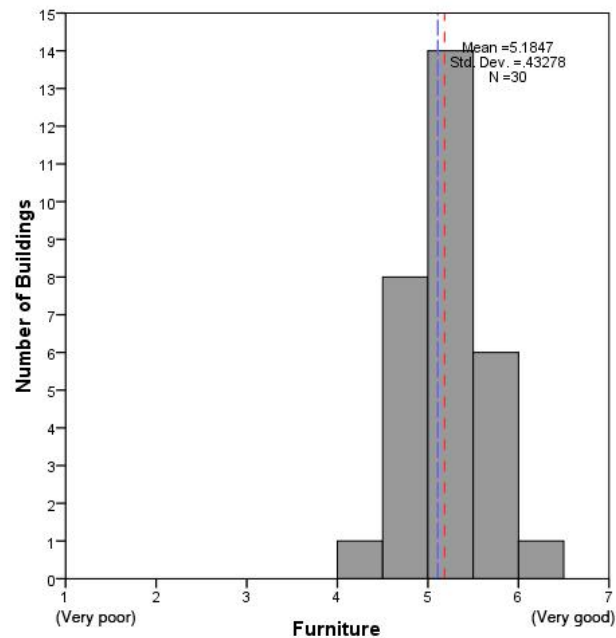


Figure 43: Frequency distribution of mean Furniture scores

Table 21: Descriptive statistics of Furniture scores

Furniture				
Sample Size (N)	Mean	SD	Skewness	Kurtosis
30	5.185	0.433	0.300	-0.194

Furniture is relatively high scoring in this sample. All of the buildings score above the scale midpoint and the mean is 5.185, so towards the satisfactory end of the scale. The kurtosis, although still negative in value is much closer to the zero and this is reflected in its more peaked distribution. The standard deviation of the distribution is also quite a lot less than some of the other questions. The occupants find the furniture in every building to be satisfactory, with the mean of one being close to the ideal score of seven.

Temperature Overall in Winter

This is the first question that the occupants are faced with that deals with the indoor environment and related issues. The occupants are asked to rate overall how they find the temperature in winter, or if they have not been in the building during a winter to ignore it. There are several questions under this heading that deal with the specific issues but this one asks for an overall assessment.

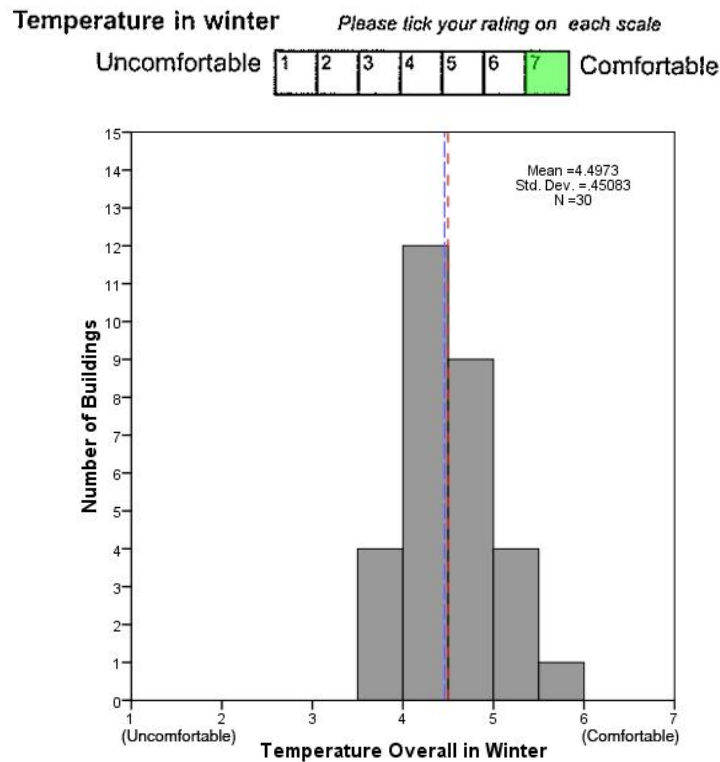


Figure 44: Frequency distribution of mean *Temperature Overall in Winter* scores

Table 22: Descriptive statistics of *Temperature Overall in Winter* scores

Temperature Overall in Winter				
Sample Size (N)	Mean	SD	Skewness	Kurtosis
30	4.497	0.451	0.360	-0.258

There is a very obvious peak, around the mean of 4.497. Although the kurtosis is a lot smaller than a lot of other variables (apart from *Furniture*), it is still negative. There is a skewness towards the lower values in the distribution. The buildings seem to be performing relatively well in this variable, with only four buildings scoring below the midpoint.

Temperature Overall in Summer

This question is phrased the same as *Temperature Overall in Winter* but is in regards to the temperature in summer.

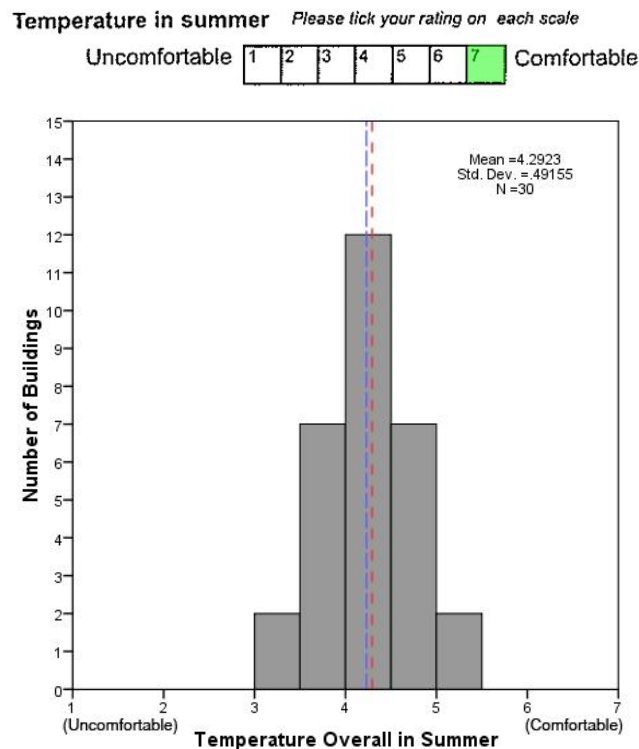


Figure 45: Frequency distribution of mean *Temperature Overall in Summer* scores

Table 23: Descriptive statistics of *Temperature Overall in Summer* scores

Temperature Overall in Summer				
Sample Size (N)	Mean	SD	Skewness	Kurtosis
30	4.292	0.492	0.182	-0.390

The main difference between this and the equivalent for winter is the less skewed distribution. Although the distribution of scores for this question and the distribution of scores for *Temperature Overall in Winter* have a similar peak in the middle, the kurtosis from this question reflects an overall less peaked distribution. This could largely be to do with the fact that the kurtosis is more influenced by the tails of the distribution than scores in the centre of the distribution (DeCarlo, 1997).

The mean is also lower indicating that this group of building performs better in winter rather than summer. There are more buildings rating below the midpoint for conditions in summer, with nine buildings compared to winter's four.

Air Overall in Winter

The last question in this series is a summary of the overall air conditions in winter.

Conditions in winter

Unsatisfactory overall ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☒ 7 Satisfactory overall

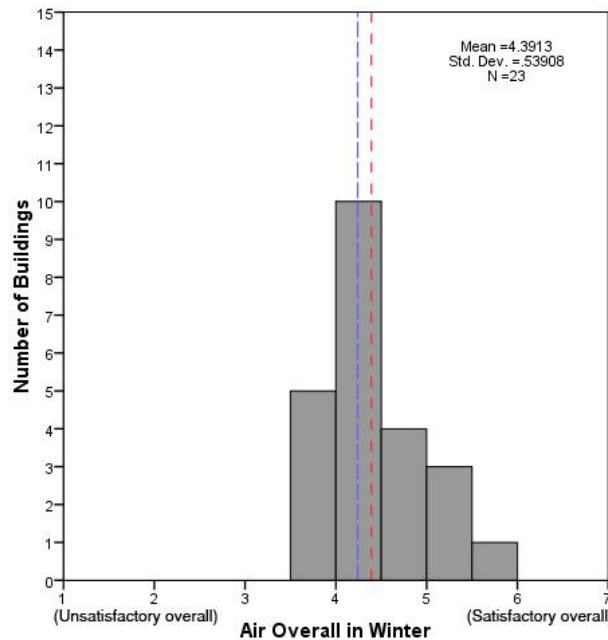


Figure 46: Frequency distribution of mean *Air Overall in Winter* scores

Table 24: Descriptive statistics of *Air Overall in Winter* scores

Air Overall in Winter				
Sample Size (N)	Mean	SD	Skewness	Kurtosis
23	4.391	0.539	0.710	0.036

The kurtosis for this question is the closest to the perfect normal distribution score of zero. However, it does have a large skewness value, which is reflected in the peaks being on the edge of the distribution towards the lower values. Five buildings are rating below the midpoint of the scale, whereas the peak in the distribution is just above the midpoint of the scale.

Air Overall in Summer

This is the corresponding question for the air conditions in summer.

Conditions in summer

Unsatisfactory overall 1 2 3 4 5 6 7 Satisfactory overall

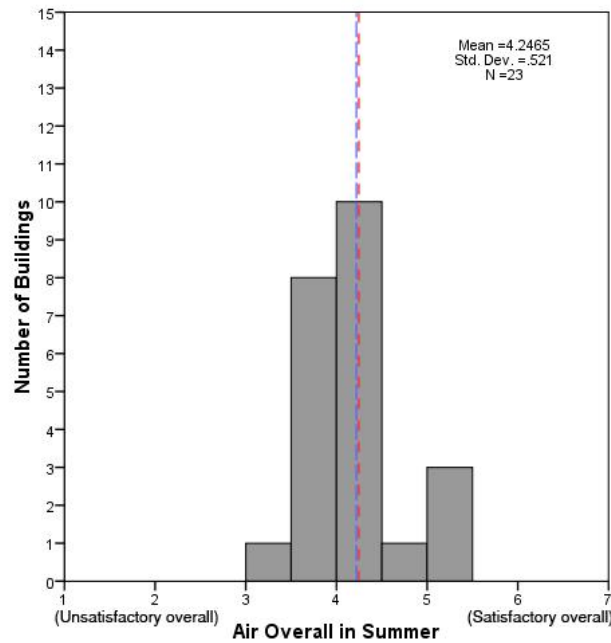


Figure 47: Frequency distribution of mean Air Overall in Summer scores

Table 25: Descriptive statistics for Air Overall in Summer scores

Air Overall in Summer				
Sample Size (N)	Mean	SD	Skewness	Kurtosis
23	4.247	0.521	0.693	0.203

These two distributions in relation to the air conditions in the building are quite similar in terms of their skewness and kurtosis values. This distribution is also skewed towards the lower values, and the kurtosis is positive. Similar to the results from temperature in winter and summer, the occupants are generally scoring slightly higher for the conditions in winter.

Noise Overall

This question in the POE questionnaire is structured as below, with the ideal answer highlighted in green. The intention is for the occupants to state how they find the overall noise level of the building, taking into account the internal and external noise. There are five sub questions under this dealing with particularly kinds of noise.

Noise overall Unsatisfactory 1 2 3 4 5 6 7 Satisfactory

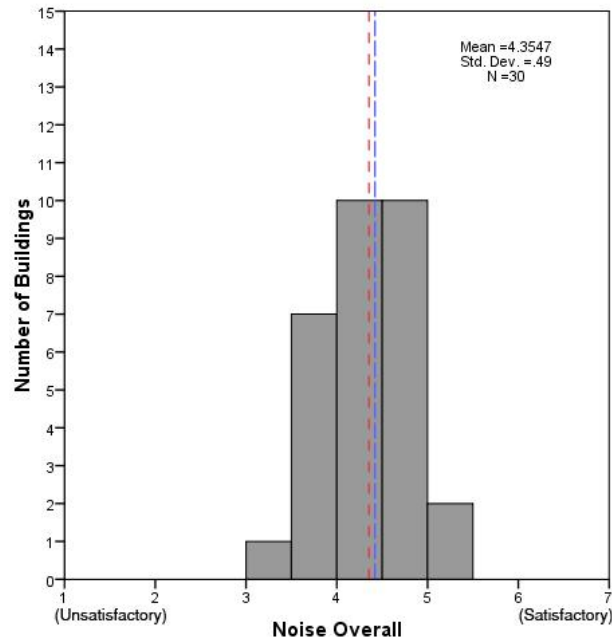


Figure 48: Frequency distribution of mean Noise Overall scores

Table 26: Descriptive statistics of Noise Overall scores

Noise Overall				
Sample Size (N)	Mean	SD	Skewness	Kurtosis
30	4.355	0.490	-0.113	-0.516

Figure 48 displays the frequency distribution of the average scores for each building. The mean value for this group of buildings is 4.355, and is highlighted with the red dashed line. This is slightly above the midpoint of the scale, so the occupants judge the buildings to be on average satisfactory in relation to noise. The scores are clustered slightly above the mid-line in the higher scoring region, resulting in the negative skewness, and the distribution is considered to be relatively flat resulting in the negative kurtosis.

Lighting Overall

This is the equivalent question dealing with lighting.

Lighting overall Unsatisfactory 1 2 3 4 5 6 7 Satisfactory

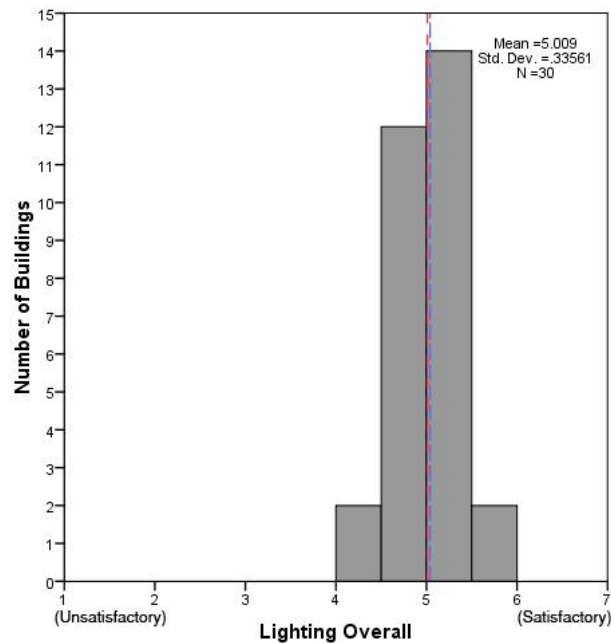


Figure 49: Frequency distribution of mean *Lighting Overall* scores

Table 27: Descriptive statistics of *Lighting Overall* scores

Lighting Overall				
Sample Size (N)	Mean	SD	Skewness	Kurtosis
30	5.009	0.336	-0.386	1.352

Simply by looking at this distribution it is remarkably different from the others, such as *Noise Overall*. There is a very obvious peak in the distribution with the adjacent bars being seven times less in number of buildings. All of the buildings are averaging scores above the midpoint of the scale. It is the highest performing overall out of the environmental factors and the second highest so far behind *Furniture*. The standard deviation is by far the smallest as almost all of the scores are clustered closely around the mean.

Overall Comfort

There are a few questions in the questionnaire that attempt to encapsulate overall performance of the building. One of these is comfort overall where the occupants are asked 'All things considered, how do you rate the overall comfort of the building environment?' This ranges from unsatisfactory to satisfactory with an ideal score of 7 and a breakeven point of 4.

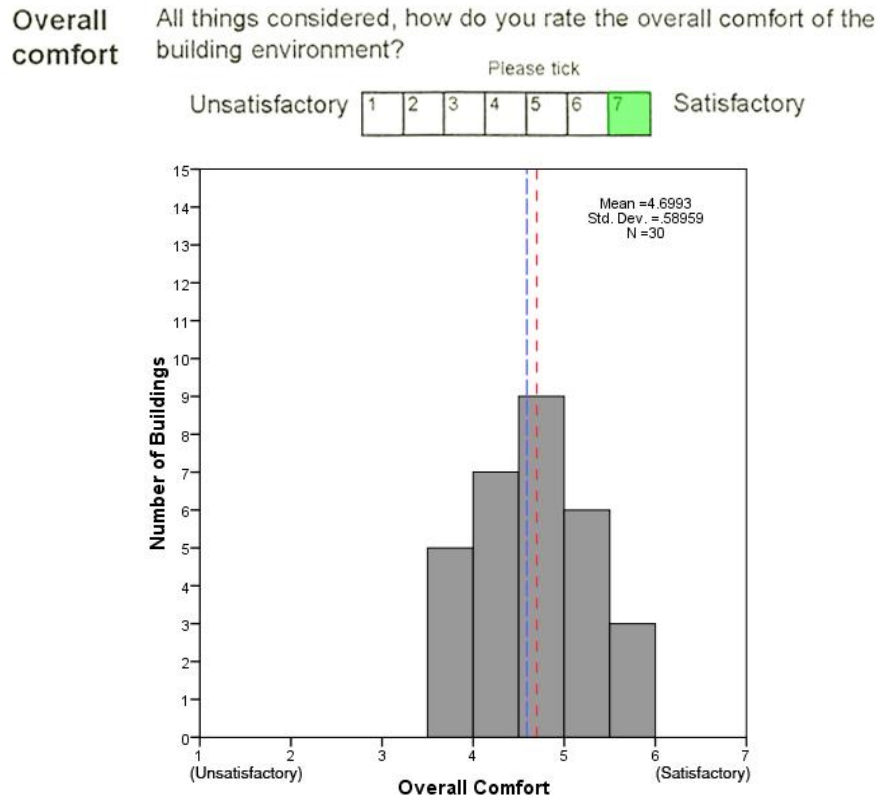


Figure 50: Frequency distribution of mean *Overall Comfort* scores

Table 28: Descriptive statistics of *Overall Comfort* scores

Overall Comfort				
Sample Size (N)	Mean	SD	Skewness	Kurtosis
30	4.699	0.590	0.213	-0.675

This is a flatter, more spread out distribution, indicated by both the larger standard deviation and the more negative kurtosis. Within the range of distribution the scores are skewed towards the lower end. All but five buildings score above the midpoint, resulting in a reasonably high mean of 4.699.

Perceived Productivity

The question on productivity is different in the fact that it utilises a 9-point scale, with each point representing a 10% band of increase or decrease in perceived productivity. The occupants are asked to rate how they perceive the building affects their productivity at work.

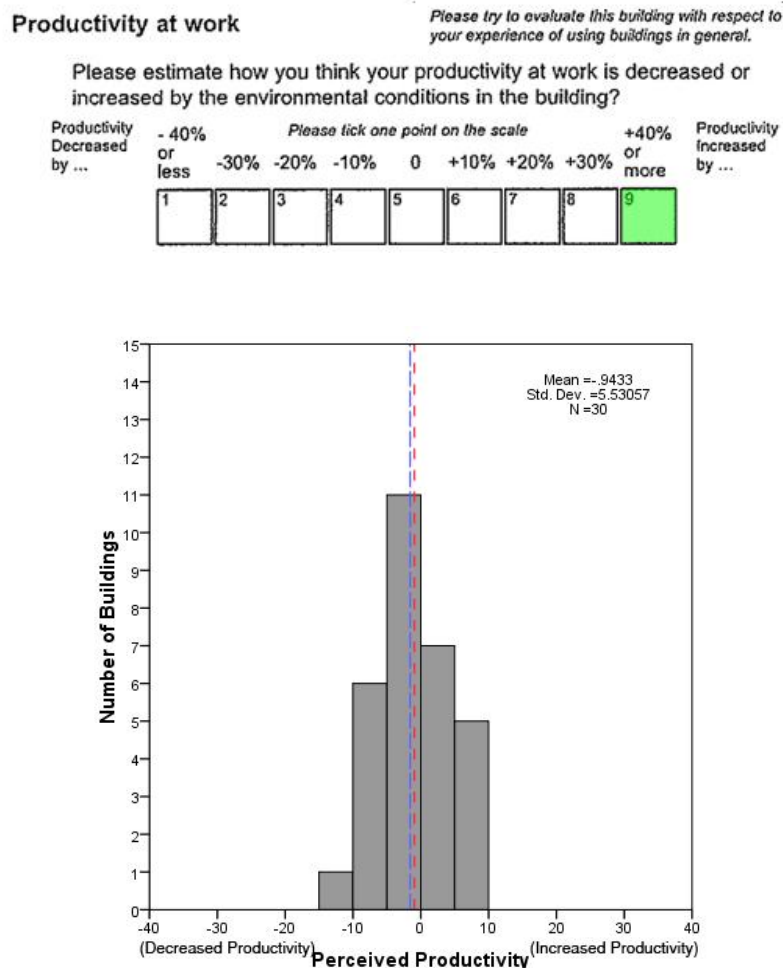


Figure 51: Frequency distribution of mean Perceived Productivity scores

Table 29: Descriptive statistics of Perceived Productivity scores

Perceived Productivity				
Sample Size (N)	Mean	SD	Skewness	Kurtosis
30	-0.943	5.531	0.277	-0.669

This group of building has a mean of decreasing productivity by 1%, therefore very close to the mid and breakeven point of the scale. Thus the indoor environment in this sample of buildings is on average negatively affecting the productivity of the occupants. In reference back to the previous section **Section 5.2: Comparison of benchmarks and databases** it can be seen that although this mean score is in the negative values it is still higher than the mean scores of previous BUS benchmarks. The mean is difficult to compare to the other questions, as it is based on percentage intervals on a 9-point scale, rather than numerical intervals on a 7-point scale. The skewness and kurtosis are however comparable. The negative kurtosis reflects the relatively flat distribution, while the positive skewness indicates that the scores in the distribution are skewed towards the lower, negative scores.

Health

Health

Do you feel less or more healthy when you are in the building?

Please try to evaluate this building with respect to your experience of using buildings in general.

Please tick

Less healthy 1 2 3 4 5 6 7 More healthy

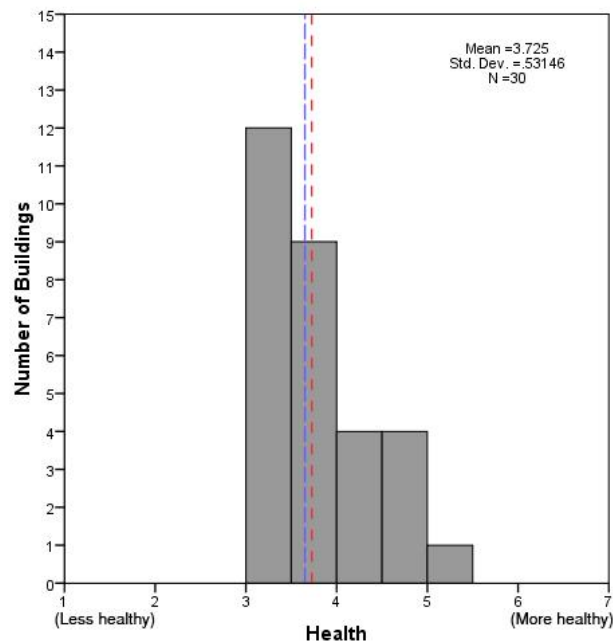


Figure 52: Frequency distribution of mean Health scores

Table 30: Descriptive statistics of Health scores

Health				
Sample Size (N)	Mean	SD	Skewness	Kurtosis
30	3.725	0.531	0.785	-0.170

This is a rather interesting distribution as on first visual assessment it does not seem to fit with a normal distribution; this will be assessed in the next section. The peak is on the edge of the distribution rather than near the middle like the others. This is particularly apparent in the high skewness value of 0.79. This is also one of two variables, the other being *Perceived Productivity*, investigated that has a mean below the midpoint of the scale indicating the occupants do not consider this group of buildings as being particularly healthy. The distribution is almost a step down pattern, going from nine to four to one building. The majority of the buildings, all but one, are still scoring in the middle range of the scale from three to five.

5.3.2 Skewness

As described in **Chapter 4.0: Research Methods**, skewness relates to the level of asymmetry in a distribution, where no skewness would be a perfect normal distribution. As this sample of buildings is on the small side in terms of sample size there is more possibility of getting non-normal data, so it is important to check whether the data is normally distributed.

If the sample is reasonably large, skewness will not ‘make a substantive difference in the analysis’ (Tabachnick & Fidell, 2001, p. 74). The risk of an underestimate of variance, which can result from kurtosis, is also reduced by a large sample (200+ cases) (Tabachnick & Fidell, 2001).

A commonly used test to assess the significance of the skewness is to check whether the skewness value calculated is more than two standard errors of skew. The standard error of skew was calculated through SPSS 16.0, and is dependent on the sample size.

Table 31: Normality test in terms of skewness

	Skewness	2x Standard Errors of Skew	Significant
Design	-0.09	0.85	No
Needs	-0.33	0.85	No
Space	0.28	1.00 ¹	No
Image	0.01	0.85	No
Meeting Rooms	-0.01	0.85	No
Storage	0.63	0.85	No
Furniture	0.30	0.85	No
Temperature in Winter	0.36	0.85	No
Temperature in Summer	0.18	0.85	No
Air in Winter	0.71	0.96 ²	No
Air in Summer	0.69	0.96 ³	No
Noise Overall	-0.11	0.85	No
Lighting Overall	-0.39	0.85	No
Overall Comfort	0.21	0.85	No
Perceived Productivity	0.28	0.85	No
Health	0.79	0.85	No

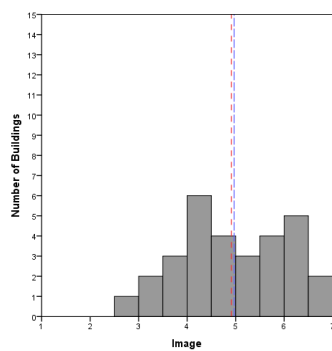


Figure 53: Smallest level of skewness

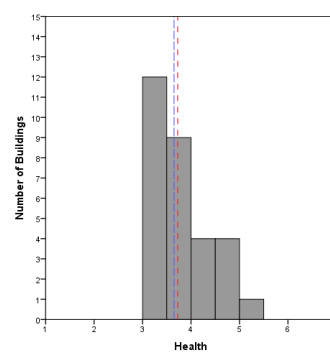


Figure 54: Highest level of skewness

¹ Results from 21 buildings, rather than the full sample of 30 buildings

² Results from 23 buildings, rather than the full sample of 30 buildings.

³ Results from 23 buildings, rather than the full sample of 30 buildings.

Therefore the level of skewness is not considered significant and it can be assumed in terms of skewness that the distribution is normal. The highest level of skewness is for *Health*, which was displayed in the earlier graph (Figure 52). Although this is relatively close to the limit of two standard errors of skew, it still can be considered to approximate a normal distribution. The least amount of skewness is in *Image* and *Meeting Rooms*, with values very close to zero.

5.3.3 Kurtosis

Kurtosis deals with the sharpness of a distribution's curve, with a perfect normal distribution having a kurtosis score of zero. 'There will almost always be a non-zero value returned, so it is useful to know when the kurtosis is significant Kurtosis is deemed to be significant when the kurtosis value supplied by the software is greater than two standard errors of kurtosis' (Bell, 2010).

Table 32: Normality test in terms of kurtosis

	Kurtosis	2x Standard Errors of Kurtosis	Significant
Design	-0.74	1.67	No
Needs	-0.34	1.67	No
Space	-1.28	1.94 ⁴	No
Image	-1.06	1.67	No
Meeting Rooms	-0.94	1.67	No
Storage	-0.72	1.67	No
Furniture	-0.19	1.67	No
Temperature in Winter	-0.26	1.67	No
Temperature in Summer	-0.39	1.67	No
Air in Winter	0.04	1.87 ⁵	No
Air in Summer	0.20	1.87 ⁶	No
Noise Overall	-0.52	1.67	No
Lighting Overall	1.35	1.67	No
Overall Comfort	-0.68	1.67	No
Perceived Productivity	-0.67	1.67	No
Health	-0.17	1.67	No

Therefore the kurtosis is not considered significant and it can be assumed in the tests, in terms of kurtosis that the distribution is normal. The two with the scores furthest from zero are for *Image* with its particularly negative score and *Lighting Overall* with its particularly positive score (from the variables with the full number of buildings). This was also displayed graphically in the histograms showed earlier (Figures 40 and 49). *Image* had a spread-out score, with no particular most common score, whereas *Lighting Overall* had a very prominent peak in the distribution.

⁴ Results from 21 buildings, rather than the full sample of 30 buildings

⁵ Results from 23 buildings, rather than the full sample of 30 buildings

⁶ Results from 23 buildings, rather than the full sample of 30 buildings

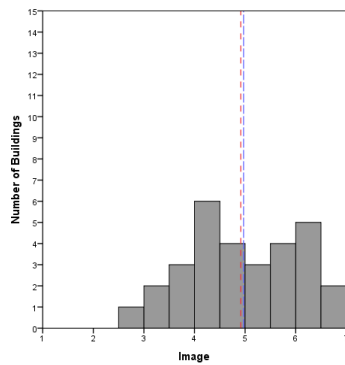


Figure 55: Negative kurtosis

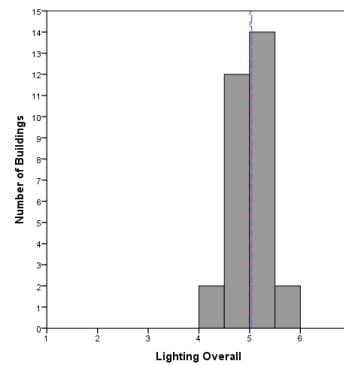


Figure 56: Positive kurtosis

5.3.4 Normal distribution test

An additional test of normal distribution was carried out using SPSS 16.0. A 'Kolmogorov- Smirnov' test was conducted to determine whether the distributions of the scores can be considered to be normal. This test 'takes the observed cumulative distribution of scores and compares them to the theoretical cumulative distribution for a normally distributed population' (Hinton, Brownlow, McMurray, & Cozens, 2004, p. 30). A sig. value above 0.05 indicates normality. This was completed over all of the questions looked at above and the results summarised in Table 33 below (full results can be found in the Appendix D).

Table 33: Kolmogorov-Smirnov test of normality

	K-S sig value	Normal/not (above 0.05)
Design	0.854	Normal
Needs	0.942	Normal
Space	0.782	Normal
Image	0.744	Normal
Meeting Rooms	0.765	Normal
Storage	0.934	Normal
Furniture	0.737	Normal
Temperature in Winter	0.999	Normal
Temperature in Summer	0.920	Normal
Air in Winter	0.816	Normal
Air in Summer	0.542	Normal
Noise Overall	0.912	Normal
Lighting Overall	0.569	Normal
Overall Comfort	0.883	Normal
Perceived Productivity	0.901	Normal
Health	0.372	Normal

The results from this test follow on from what was shown in the examination of the skewness and kurtosis tests. The result of all these scores being normally distributed means that any of the appropriate statistical tests can be carried out, and the statistics already looked at in the previous section can be applied, such as the standard deviation.

They are comparable to the other benchmarks looked at in this research, so the same techniques are able to be used without violating the assumption of normality.

5.3.5 Summary

In terms of means of the scores *Furniture* received the highest score with 5.185, followed by *Lighting Overall* with 5.009. The lowest scoring variables were *Health* with 3.725, the only mean score below four, followed by *Storage* with 4.090. *Image* has the largest variability with a standard deviation of 1.132, while *Lighting Overall* has the smallest variability with a standard deviation of 0.336.

Perceived Productivity is in its own category due to the different scale it uses for comparing means and standard deviations. The skewness and kurtosis are comparable as they look at the difference from normal rather than the scale that it is based on.

Five of the variables have a negative skewness, while the remaining eleven have a positive skewness and are therefore skewed towards the lower values. The distributions are almost all negative in kurtosis, meaning that the distributions are relatively flat, with the exception of *Lighting Overall*, *Air Overall in Summer* and *Air Overall in Winter*. These are small enough to not have a significant effect, and for the distribution to approximate normality.

An argument of Cohen et al (2006) in **Section 3.2.1: Energy benchmarks** was that energy data is more likely to be positively skewed and that this means the median is the more appropriate measure of central tendency to use. From this questionnaire data, the fact that the distribution of all of these questions can be considered normally distributed results in the mean being able to be used, as there is not the same level of skewness in this data as in Cohen's.

'Averages are more meaningful when the responses to a rating scale show a normal distribution' (Oseland, 2007, p. 37)

This result of the variables being normally distributed could largely be to do with the central limit theorem. This is when 'for fairly large samples, the distribution of sample means has been shown to reasonably approximate a normal distribution even if the individual values are distributed very differently' (Aronoff & Kaplan, 1995, p. 353).

Parametric tests are able to be carried out on these distributions without an underlying assumption being violated.

'The existence of flat or peaked distributions as indicated by the kurtosis statistic is important to you as a language tester insofar as it indicates violations of the assumption of normality that underlies many of the other statistics like correlation coefficients, t-tests, etc. used to study the validity of a test. Skewed distributions will also create problems insofar as they indicate violations of the assumption of normality that underlies many of the other statistics like correlation coefficients, t-tests, etc. used to study test validity' (Brown, 1997).

These tests will be carried out in the next section. Correlation coefficients will be produced to assess the relationship between the variables, followed by the independent t-test to compare the different building variables.

5.4 Questionnaire variables

5.4.1 Correlations

Similar studies to this one exploring the nature of questionnaires, have demonstrated that there are strong correlations between variables in the questionnaire being investigated (Leaman & Bordass, 2007; Leaman, Thomas, & Vandenberg, 2007; Baird & Oosterhoff, 2008; Baird, 2009). The 16 summary variables with an ideal score of seven were compared to each other to see if there was a correlation between how the respondents were scoring the different variables, and to see how, if at all, the variables related to each other.

In this sample, there were multiple correlations that were considered to be 'very strong' and 'strong' indicating that a lot of the variables are closely related to each other. For a full list of the individual correlations refer to Appendix E.

The graph below displays pictorially the highest correlation that was found in the sample, between *Design* and *Image*. It plots the mean scores from this group of buildings for *Design* against the mean scores for *Image*. The line of best fit and 95% confidence levels are also displayed.

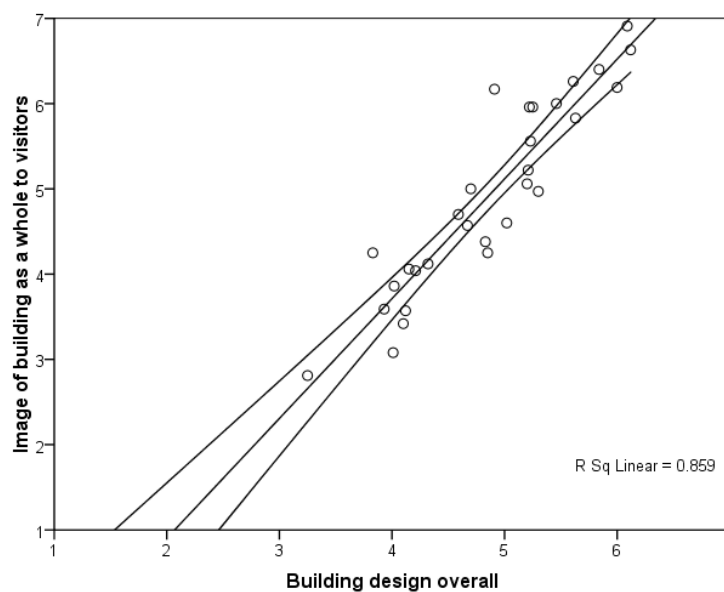


Figure 57: Highest correlation in sample - *Design* and *Image*

The R squared value gives an indication of the strength of the relationship between the two. This R squared linear value of 0.859 (being the square of the correlation coefficient) signifies that it accounts for around 86% of the variance between the two variables. The relationship is considered to be 'very strong' (Salkind, 2005). There is a positive relationship, so generally the higher people score for *Design* the higher they rate the building for *Image* and vice versa.

In keeping with the aspects previously identified that have an effect on the occupants of a building (**Section 2.1: How the building environment affects the health, comfort, and productivity of building occupants**), the correlations with these three variables were looked at in further detail to determine what variables have the biggest impact on them. Table 34 below shows the correlations between the summary variables and these three specific variables and their corresponding Pearson correlation coefficients.

Table 34: Correlations between summary variables

	Health	Comfort	Productivity
Design	0.680	0.921	0.814
Needs	0.645	0.888	0.836
Space	0.637	0.824	0.781
Image	0.598	0.806	0.709
Meeting Rooms	0.441	0.726	0.696
Storage	0.606	0.854	0.818
Furniture	0.562	0.782	0.812
Temperature in Winter	0.635	0.872	0.763
Temperature in Summer	0.661	0.816	0.760
Air in Winter	0.680	0.796	0.782
Air in Summer	0.746	0.750	0.736
Noise Overall	0.428	0.693	0.708
Lighting Overall	0.438	0.627	0.707
Overall Comfort	0.737	-	0.902
Perceived Productivity	0.713	0.902	-
Health	-	0.737	0.713

Using Salkind's rule of thumb for determining the strength of relationships between variables, there are multiple 'very strong' and 'strong' relationships. These show that the three variables are closely related, with *Health* a slightly less influential variable than the other two as there are no 'very strong' correlations in this dataset with these results. This was foreseen through the analysis of the distribution of scores, as *Health* had quite a different pattern of distribution from the other summary variables, and was the lowest scoring in the comparison to the established BUS benchmarks.

There is a strong relationship between most of the scores. This could be largely to do with their using the same scale (apart from *Perceived Productivity*), as well as the fact that they are based on mean scores of the buildings. It should also be noted that all of these correlations are positive. This means that as one of the variables scores higher so does the corresponding one. This indicates that if a building scores quite highly in an overall aspect this is likely to be followed through in others as the occupants generally like the building.

The highest correlation with these variables is between *Design* and *Overall Comfort*, with a Pearson coefficient of 0.921, followed by *Overall Comfort* and *Perceived Productivity*, with a Pearson coefficient of 0.902. Both of these are considered to be 'very strong' relationships. The strong relationship between *Overall Comfort* and *Perceived Productivity* is carried through in other studies (Leaman & Bordass, 2007; Leaman, Thomas, & Vandenberg, 2007; Baird, 2009).

Leaman explores the correlations between different environmental variables and *Perceived Productivity* in his set of 177 UK buildings, looking at buildings with different ventilation methods separately. He identifies 'very strong' correlations for all ventilation methods between *Overall Comfort* and *Perceived Productivity* with the lowest in the air conditioned (AC) buildings with a Pearson correlation coefficient of 0.827 and the highest in the mixed mode (MM) buildings with a Pearson correlation coefficient of 0.874 (Leaman & Bordass, 2007, p. 667).

This particular correlation has also been looked at in terms of the 2009 BUS NZ benchmark. In Baird's article he incorporates a correlation of *Overall Comfort* and *Perceived Productivity* using the 2009 BUS NZ benchmark database, to emphasise the strong relationship that is present between these two variables. The Pearson correlation coefficient that was identified in this dataset is 0.870, so is similar to the UK buildings and smaller than found in this study (Baird, 2009).

Correlations with *Health* have also been looked at in more detail in previous studies. In a study focussing on sustainable buildings the correlations with *Health* were shown to be stronger than found in this study. However, *Perceived Productivity*, *Overall Comfort* and *Design* were all the variables with the strongest relationship with *Health* (Baird & Oosterhoff, 2008).

Section 2.1: How the building environment affects the health, comfort, and productivity of building occupants determined some features of the indoor environment that can affect the occupants. In terms of *Health* one of the main affects was the air quality. From these questionnaire results, in terms of the indoor environment variables the most influential variable on *Health* is *Air Overall in Summer*, reflecting the already established relationship.

In terms of *Overall Comfort*, *Temperature Overall in Winter* has the strongest relationship, and for *Perceived Productivity* it is both the air quality and the temperature.

5.4.2 BUS summary method comparison

The relationship between the variables is also explored through the summary methods that BUS produce in their report. These are the four indices, and the BUS rating score: selected variable method. These were calculated for the pilot database buildings using the BUS UK 2006 database, and the BUS NZ 2008 database for the indices, due to the availability of data, and the BUS NZ 2010 for the rating scores.

Indices

As described in **Section 3.4: Assessment of BUS benchmarks** there are three indices that are created to get an idea of the overall scoring of the occupants for the buildings. These are the *Comfort Index*, *Satisfaction Index* and *Summary Index*. As these are used to encapsulate an overview of occupants' perceptions of several performance variables, they can be looked at as an indication of the overall performance of the buildings. There is also the *Forgiveness Factor* that quantifies the tolerance that the occupants have for the buildings.

The indices in the reports were created using different benchmark data, over time and over locations. To allow a comparison to a constant the BUS UK 2006 dataset was used as a basis for the benchmark mean and standard deviations that are used in the equations. Complete questionnaire data was not available in all cases resulting in the *Comfort Index* not being able to be calculated for some, hence the smaller number of results.

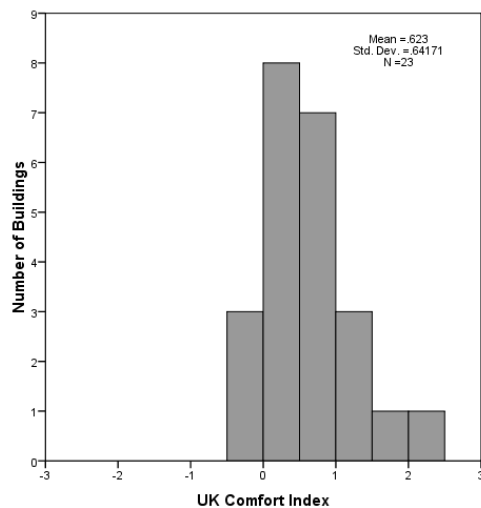


Figure 58: UK *Comfort Index* for pilot database buildings

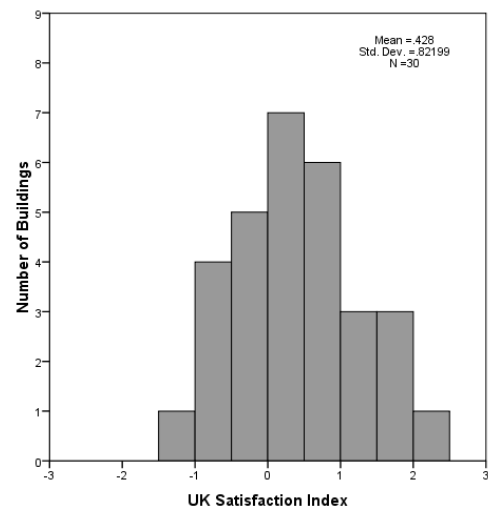


Figure 59: UK *Satisfaction Index* for pilot database buildings

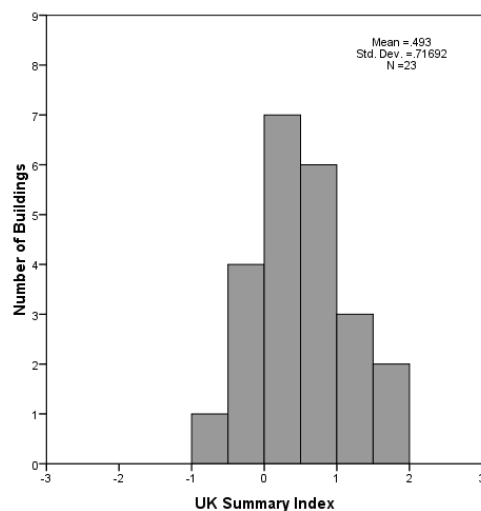


Figure 60: UK *Summary Index* for pilot database buildings

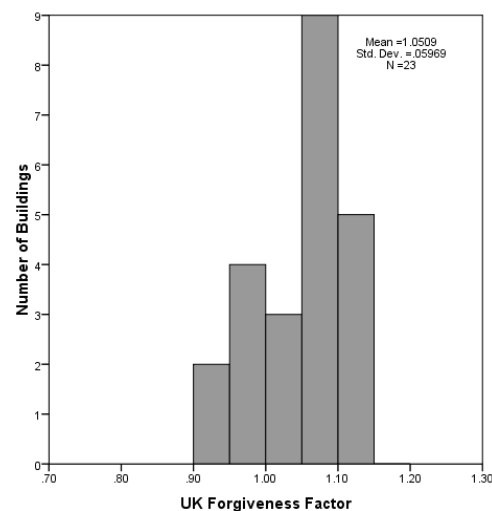


Figure 61: UK *Forgiveness Factor* of pilot database buildings

For the *Comfort Index*, as shown in Figure 58, 20 of the 23 cases were above the scale midpoint and five were greater than +1.00. Only three were below the midpoint, and none of these are below -1.00.

The *Satisfaction Index*, shown in Figure 59, was able to be calculated for all 30 buildings. From these results 20 out of the 30 cases were above the scale midpoint of zero in terms of satisfaction. 10 buildings are below the midpoint and from these one is below -1.00.

The *Summary Index* is the mean of the *Comfort* and *Satisfaction Indices* and this is reflected in its distribution. The scores range from -1.00 to +2.00, with a mean score of 0.50. 18 cases are above the midpoint of the scale and four are below.

This same method was carried out to look at the buildings in relation to the BUS 2008 NZ benchmark. The resultant graphs are shown below.

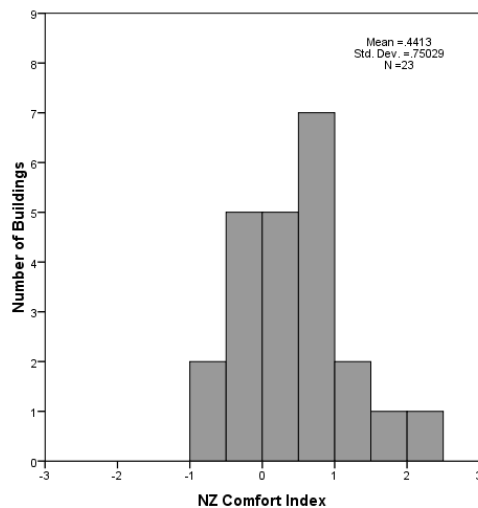


Figure 62: NZ *Comfort Index* for pilot database buildings

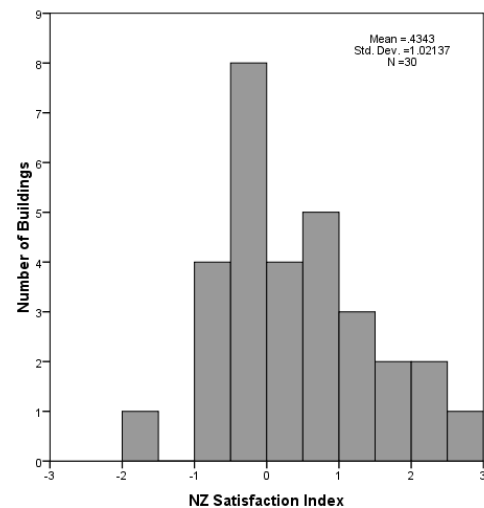


Figure 63: NZ *Satisfaction Index* for pilot database buildings

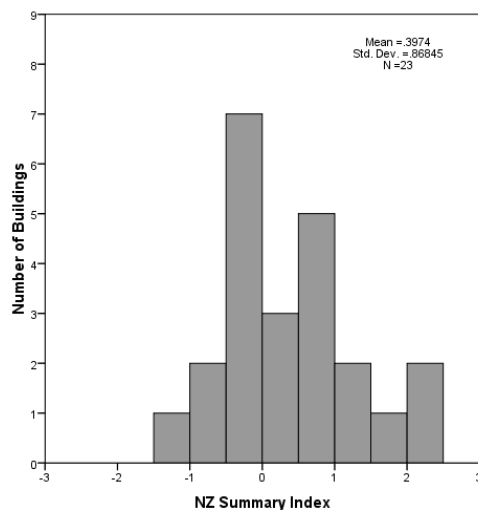


Figure 64: NZ *Summary Index* for pilot database buildings

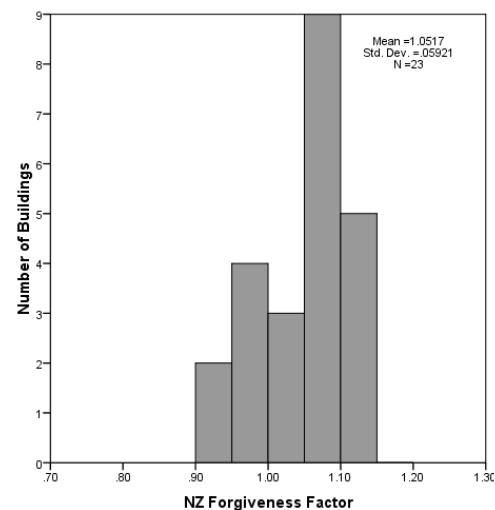


Figure 65: NZ *Forgiveness Factor* for pilot database buildings

Using the NZ benchmark the indices have a wider range over the buildings. The *Satisfaction Index* for example has 13 buildings scoring less than zero, which is more than for the UK comparison, but also has a higher scoring building, just under +3.00.

The mean for the *Comfort Index* in the UK assessment is significantly higher than this NZ one. Using the NZ benchmark the means for the *Comfort* and *Satisfaction Indices* are much closer to each other.

The *Forgiveness Factors* are constant over both assessments as they do not rely on a benchmark score for a comparison. All but six buildings have *Forgiveness Factors* greater than one, indicating that the occupants are quite tolerant of the environmental conditions. This is an indication of how the occupants are more likely to rate the overall scores higher, so the overall scores in the pilot database may be different than the more direct questions. For example scores for *Noise Overall* may be higher than the scores for more specific questions such as *Noise from Colleagues*. Tolerance of unsatisfactory conditions may be influenced by the type of building owner or employer; if they are interested enough in performance of their building to be doing such surveys or trying to improve conditions, they may be seen as good employers and therefore produce happier building occupants.

Baird (2010) has looked at how the individual variables relate to these indices for international sustainable buildings. Correlations between the *Comfort Index* and the variables that make it up were ranked as first *Overall Comfort* (0.875), *Air Overall in Summer* (0.822), *Temperature Overall in Winter* (0.806), *Air Overall in Winter* (0.796), *Noise Overall* (0.783), *Temperature Overall in Summer* (0.748), and *Lighting Overall* (0.562). For the *Satisfaction Index* ranking order was *Perceived Productivity* (0.954), *Design* (0.913), *Health* (0.912), and *Needs* (0.888), all in the very strong category.

Carrying out this same analysis on the data from this set of 30 NZ buildings gave slightly different results. For the *Comfort Index* the ranking order of the correlations was the same with both the UK benchmarks and the NZ benchmarks, which was similar but not identical to Baird's international sustainable set. The variables were ranked *Overall Comfort*, *Temperature Overall in Winter*, *Air Overall in Summer*, *Temperature Overall in Summer*, *Air Overall in Winter*, *Lighting Overall* and *Noise Overall*. For actual figures refer to Appendix E. For the *Satisfaction Index* the ranking order was slightly different between the two benchmarks. For the correlation with the *Satisfaction Index* calculated with the BUS UK benchmark the ranking order was *Design*, *Perceived Productivity*, *Needs* and *Health*. For the correlation with the *Satisfaction Index* using the BUS NZ benchmark the ranking order was *Design*, *Needs*, *Perceived Productivity* and *Health*, all in the 'very strong' category.

The correlations were also found between the *Summary Index* and the 11 summary variables that make up both the *Comfort Index* and the *Summary Index*. The ranking order of the correlations was consistent in both the UK and NZ benchmark comparisons. The variables were ranked, from strongest to weakest correlations, *Overall Comfort*, *Perceived Productivity*, *Needs*, *Design*, *Health*, *Temperature Overall in Winter*, *Air Overall in Winter*, *Temperature Overall in Summer*, *Air Overall in Summer*; all rating in the 'very strong' category, followed by *Noise Overall* and *Lighting Overall*; in the 'strong' category.

From these correlations it can be seen that the *Temperature* and *Air* variables seem to be the most influential out of the indoor environmental variables, with *Noise Overall* and *Lighting Overall* being the least. This is consistent with the findings of the variable correlations. In terms of the correlations with the *Summary Index*, the satisfaction variables; *Overall Comfort*, *Perceived Productivity*, *Needs*, and *Design*, have the strongest relationship. *Overall Comfort* in particular correlates strongly with all three of the summary indices.

Rating scores

The benchmark datasets, UK 2006 and NZ 2010, were then used to analyse the buildings in terms of the rating scores. The BUS rating score: selected variable method was used and the method followed was that used by BUS and described earlier in **Section 3.4: Assessment of BUS benchmarks**. The result of the rating score method is a score out of seven that can be used as an overall rating of the performance of a building from the users' point of view.

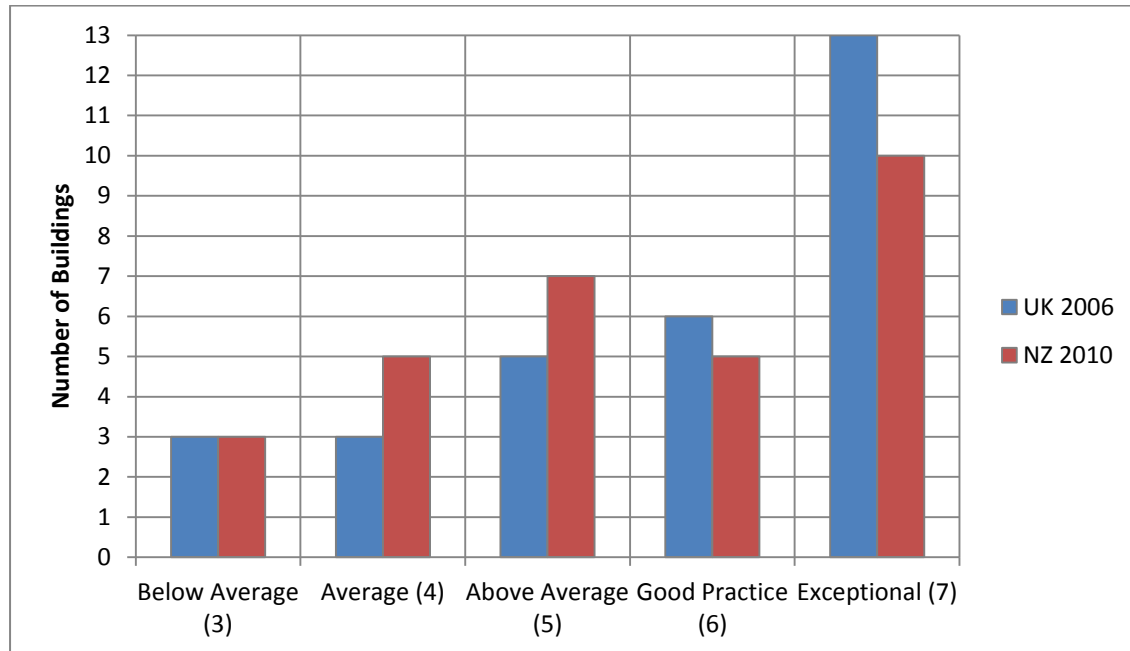


Figure 66: BUS rating score: selected variable method results for pilot database buildings when compared to BUS benchmarks

A large majority of these buildings performed very well, as indicated by the 7/7 ratings of 13 buildings compared to the UK benchmark, and 10 buildings compared to the NZ benchmark. It follows a similar pattern to what was shown in the indices, with more buildings scoring towards the higher values than the lower. As seen in **Section 5.2: Comparison of benchmarks and databases** the BUS NZ 2010 benchmark had a consistently larger range of values within the confidence limits. This could be why the comparison with the NZ benchmark resulted in more buildings rating in the middle of the range, as they are rating similarly to the benchmark and scale midpoint.

In 'Sustainable Buildings in Practice' (Baird, 2010) the correlations were looked at between the BUS rating score: selected variable method and the ten variables in the questionnaire that are used to calculate this. This is similar to the method carried out for the different indices, above. Baird (2010) found that the correlations between the rating on a 7-point scale and the rating on a percentage scale were very strong, with a Pearson correlation coefficient of 0.984, so could be considered justification for using either. For the set of NZ buildings used in this project, the relationship for the buildings and rating score using the NZ 2010 benchmark and corresponding percentage score was 0.976 and that from the UK 2006 benchmark 0.977, so both are slightly lower than that for the sustainable buildings but these are still very strong correlations.

The ranking of the correlations were relatively consistent for both the UK rating score and the NZ rating score. *Overall Comfort* had the strongest relationship for both benchmarks. This was followed by, for UK rating score; *Design, Needs, Temperature Overall in Winter, Perceived Productivity, Image, Temperature Overall in Summer, Noise Overall, Health, and Lighting Overall*. For the NZ rating scale *Perceived Productivity* rated third, and *Temperature Overall in Winter* moved down to below *Image* (refer to Appendix E for actual values).

5.5 Building variables

This section focuses on differences in the results between buildings with some of the significant factors identified in **Section 5.1: Classification of design**. It will first look at the effect of the sustainable nature of the buildings, followed by the size and the ventilation method.

The variables identified in **Section 5.1: Classification of design** could be looked at with the use of ANOVA (a procedure for determining whether significant differences exist between two or more sample means) or independent t-tests to try and determine whether there are significant differences between building features.

Often there are splits requested in other databases when the client wants to focus on a particular aspect. In the CBE database:

'Frequent benchmarking requests include comparisons within a single building type (hospitals, schools, etc.) within a time frame (e.g. 1998-2008) within a region (the tri-county metropolitan Detroit area) or among buildings that include the same feature (e.g. operable windows of under floor air distribution)' (Center for the Built Environment, 2011, p. 9).

Design Intent

One particular area of building design for which occupant questionnaires are undertaken are buildings that have some design intent to be sustainable. These buildings are designed to have an indoor environment that is particularly pleasant for the occupants, as well as minimising the use of energy and resources. This has been looked at in many studies (Leaman & Bordass, 2007; Leaman, Thomas, & Vandenberg, 2007; Baird, 2010; Baird, Leaman, & Thompson, in press). One such is Baird's (2010) study of sustainable buildings from around the world and their occupant questionnaire rating. The framework for the inclusion of these buildings in the study is previously stated in **Section 4.2.1: Previously surveyed buildings**. Another is the Health Optimisation Protocol for Energy-efficient Buildings (HOPE) project that selected the buildings on the basis that they included energy saving measures (Bluyssen, Aries, & van Dommelen, 2011).

The buildings from the HOPE analysis were classified into groups depending on whether they included energy saving measures or not. Whether the energy saving measures have been carried through to practice and produced a more energy efficient building has not been determined (Bluyssen, Aries, & van Dommelen, 2011).

The buildings in the pilot database were characterised into groups as to whether they were designed to be sustainable or conventional. Table 13 in **Section 5.1: Classification of design**, shows the classifications for each building in the pilot database, and Table 14 illustrates their sustainable design features. An independent t-test was carried out to determine if there was a difference in these two groups, to help determine whether there should be one benchmark, or whether sustainable buildings should be aiming higher.

An 'independent samples t-test' was conducted to explore the impact of the design intentions employed in the building on the occupants' perceptions of the buildings. The buildings were split into the two groups: Sustainable or Conventional. This was completed over the summary questions previously looked at and the results summarised in the Table 30 (full results can be found in Appendix F). The first two columns in the table relate to the test of appropriateness; if the sig. value is above 0.5 equal variances can be assumed. The second two columns relate to the actual differences between the samples and determine whether there is a difference in the way the occupants are scoring; if the sig. value is below 0.5 a significant difference can be assumed. The last two columns focus on the size of this difference, and using the equation stated in the methodology whether the difference can be considered small, medium or large.

Table 35: Independent t-test in relation to sustainability

	Hom. Of Var.		t-test		eta ²	
	sig.	equal variances assumed	sig.	difference	sig.	effect
Design	0.565	yes	0.002	yes	0.306	large
Needs	0.737	yes	0.016	yes	0.191	large
Space	0.723	yes	0.014	yes	0.279	large
Image	0.826	yes	0.000	yes	0.458	large
Meeting Rooms	0.501	yes	0.003	yes	0.273	large
Storage	0.140	yes	0.105	no	-	-
Furniture	0.191	yes	0.006	yes	0.237	large
Temperature Overall in Winter	0.354	yes	0.026	yes	0.165	large
Temperature Overall in Summer	0.385	yes	0.031	yes	0.155	large
Air Overall in Winter	0.109	yes	0.129	no	-	-
Air Overall in Summer	0.198	yes	0.258	no	-	-
Noise Overall	0.807	yes	0.102	no	-	-
Lighting Overall	0.847	yes	0.017	yes	0.187	large
Overall Comfort	0.119	yes	0.005	yes	0.252	large
Perceived Productivity	0.163	yes	0.001	yes	0.322	large
Health	0.066	no	0.018	yes	0.192	large

There is generally a difference between the buildings classed as sustainable and the conventional. All of them are significant at the 0.05 level. This could also partly be to do with the age of the building, as the sustainable buildings are almost all newer than the others. In every case the sustainable buildings rate higher than the conventional buildings.

The two variables with the largest differences between the mean scores for the two groups are *Image* and *Perceived Productivity*. It can be seen pictorially in Figures 67 and 68 that there is an obvious difference in distribution, particularly with *Image*. The occupants of the buildings classed as sustainable rate the image of the buildings higher than those in the more standard buildings.

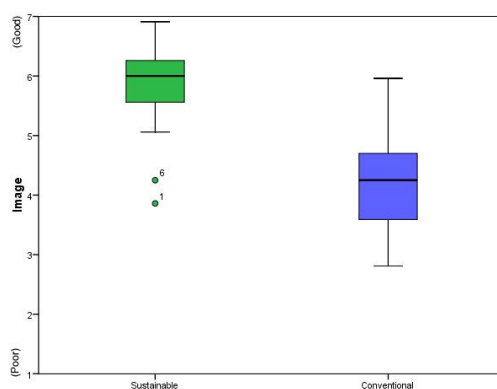


Figure 67: Large difference - *Image*

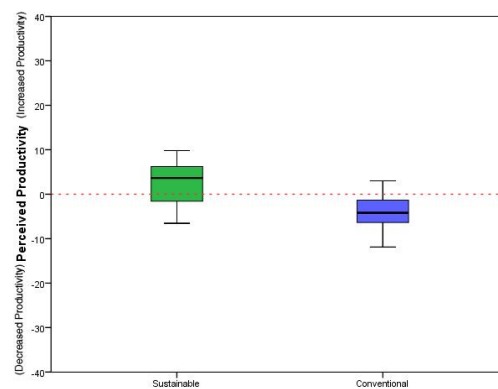


Figure 68: Large difference - *Perceived Productivity*

There are four variables where there is no significant difference between the sustainable group of buildings and the conventional. The two variables with the smallest differences with all 30 buildings are displayed pictorially below in Figures 69 and 70.

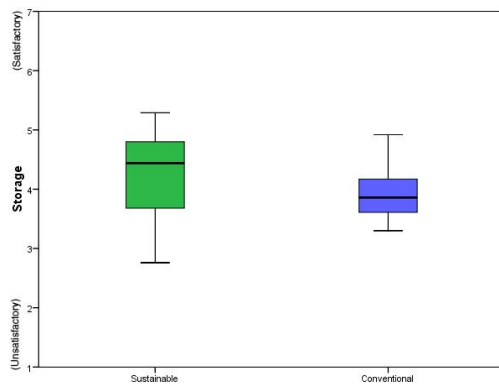


Figure 69: No significant difference - *Storage*

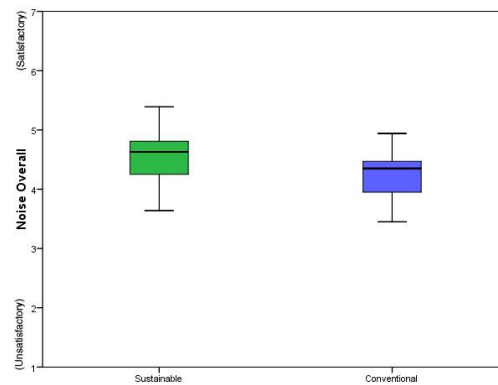


Figure 70: No significant difference - *Noise Overall*

The distribution for *Noise Overall* in the buildings is remarkably similar. While the means for *Storage* are reasonably different, the scores from the sustainable buildings are much more wide ranging than those for the conventional buildings. This is displayed numerically with the smaller homogeneity of variance sig. value in the first column in Table 35.

These two variables were also noted in **Section 5.3: Overview of scores of the buildings** to be generally low scoring, and have been considered to be the largest source of complaint in previous sets of buildings (Baird, 2010).

Baird et al (in press) expanded on this in a study that compared the differences between international sustainable and conventional buildings. 31 sustainable buildings were compared to a sample of conventional buildings ranging in number from 47 to 109 depending on the variable. The larger sample sizes allow more confidence in the results received in this study. It is reassuring to see that a lot of the results are similar. There was a significant difference between the majority of the variables. In the international building test, all of the summary variables tested above apart from *Noise Overall* had a significant difference. *Air Overall in Winter*, *Air Overall in Summer* and *Storage* were considered to be different, whereas in the NZ test the difference was not enough to be considered significant. The variation in results could be due to the smaller sample size. When a larger database is sourced this should be tested again for the differences to be fully realised. The differences could be due to the country, as the tests run in this project were only on New Zealand buildings, while Baird's was an international dataset.

Floor Area

The floor area is directly related to the number of people in the space and this could affect the ratings of the occupants, so should also be looked at, but as the floor areas are currently split into five groups there is not enough in each group to determine a difference.

As this sample is mainly made up of buildings with larger floor areas it is not feasible to assess this. This could be a test to run when the database is larger and there is a more even distribution of differing floor areas.

The majority of the questionnaire based surveys studied identified the different sizes in the buildings, but did not run tests to determine if there was a significant difference in the scores received from the occupants in them.

Ventilation

Many of the questionnaire based methods also focussed on particular ventilation methods. The building samples that the projects used were classified into ventilation categories in HOPE, SCATS, and several of the BUS studies. Adrian Leaman in his study of green buildings compared to conventional buildings split his data into three groups; NV, MM and AC (Leaman & Bordass, 2007). Through a study by de Dear and Brager it was determined that the occupants are likely to be satisfied over a wider temperature range in buildings that employ natural ventilation than those that use centrally controlled air conditioning systems (de Dear, Brager, & Cooper, 1997).

As for the pilot database used in this study, there is an uneven grouping of the three different ventilation methods, with natural ventilation being particularly unrepresented. This results in it not being feasible to carry out reliable statistical tests. The difference that the ventilation methods have on NZ buildings will be able to be confirmed with the addition of the BEES buildings, when an ANOVA can be run.

Location

The location of the building, as stated in **Section 2.1: How the building environment affects the health, comfort, and productivity of building occupants**, can have an effect on the buildings that are produced, making it a classification area that should be explored further. The pilot database is restricted to New Zealand buildings, but the distribution of these buildings around the country is not evenly spread. The large majority of the buildings are in Wellington. The splits that are made in the BEES data, and were stated in **Section 4.3.1: Classification of design**, look at the country in two groups; Auckland and the rest of NZ. There are not enough Auckland buildings in the pilot database to determine the relationship between these two groups of buildings and location.

The existing questionnaire based methods that were previously studied identified which country the buildings were from, and often used country-based averages as a comparison. This relates back to previously stated points about the importance of country-based benchmarks.

SCATS carries out a country comparison between results and determines that 'such variation between countries could also make the development of international standards for certain *individual* aspects of the indoor environment difficult, and perhaps even undesirable, at least until the nature of satisfaction in relation to language and culture, and the dynamic of the combination of its many aspects are better understood' (Humphreys, 2005, p. 324).

5.6 Analysis of scores summary

This chapter has explored the nature of the results received from the BUS questionnaire for the pilot database buildings. This included an assessment of the buildings themselves, a comparison of benchmarks, the distribution of scores and the relationship between variables.

It was determined that the buildings in the pilot database are not representative to the NZ building stock. A larger sample would be needed to make representative benchmarks. The majority of the buildings were situated in Wellington, and were of the larger size grouping.

The database means and confidence intervals were then compared to the BUS databases, which are an already established method. It was determined that the pilot database results were very similar to the existing BUS NZ 2010 benchmark and hence were valid to explore further. In the BUS method a small difference between the means of the databases can be carried through to a larger difference in the '*Better*,' '*Similar*,' '*Worse*' method. The buildings rated higher in the BUS UK 2006 benchmark comparison, than the BUS NZ 2010 and the pilot database.

The distributions of scores were quite different over the questions, but were all able to be considered normal. The majority of the buildings were scoring above the mid-point of four for most of the questions. The only questions where the means were below the mid-point were for *Health* and *Perceived Productivity*. *Health* was also the most skewed distribution, towards the low end of the scale. The highest mean scoring question was for *Furniture*, with 5.185, followed by *Lighting Overall* with 5.009.

It was determined that as the 'semantic differential' scales were able to be considered interval scales, and that the distributions could be considered to approximate normality, that parametric tests were able to be carried out on the data. This is desirable as parametric tests are generally considered more powerful than non-parametric tests.

'Statisticians prefer to treat ratings as if they were equal interval variables because more powerful parametric statistical analyses can be applied. Parametric statistics assume that the data being analysed are normally distributed. Where the distribution of sample data does not reasonably approximate a normal distribution, nonparametric statistics should be used. They are less powerful than parametric statistics (i.e., for a given sample there will be a lower level of confidence in the statistical assertions that can be made), but they do not assume that the data conform to a particular distribution' (Aronoff & Kaplan, 1995, p. 349) .

Correlations were carried out on the summary variables, with particular focus on *Health*, *Overall Comfort*, and *Perceived Productivity*. These were highly correlated with each other and many of the other variables. In terms of the indoor environment variables, *Air Overall in Summer*, *Air Overall in Winter*, *Temperature Overall in Summer* and *Temperature Overall in Winter* were consistently highly correlated with *Health*, *Overall Comfort*, and *Perceived Productivity*. *Lighting Overall* had the weakest correlation for all three variables. This was consistent with findings from the Probe analysis.

'One of the emerging findings from Probe (it has yet to be tested more fully) is that lighting – unless it is very good or very poor – has little influence on the occupants' rating of overall comfort or associated variables' (Leaman & Bordass, 2001, p. 136).

The BUS summary methods were calculated for the pilot database using the BUS NZ 2008 benchmark and the BUS UK 2006 benchmark. A large majority of the buildings were high scoring. They rated higher in the comparison with UK benchmark than the NZ benchmark.

In terms of how the building variables affect the resultant scores, several of the methods that could be considered significant were not able to be tested, due to the small sample size. An independent t-test was carried out on the pilot database, and compared to a study carried out by Baird et al (in press), in terms of the difference between sustainable and conventional buildings. There was considered to be a difference in scoring between the groups of buildings for almost all of the variables tested, which was consistent with the findings of Baird et al (in press).

6.0 USER PERCEPTION BENCHMARKS AND THEIR INCORPORATION INTO RATING TOOLS

This section explores potential approaches for the creation of benchmarks for users' perception scores and the incorporation of these scores into rating tools. This can be considered in terms of both questionnaire data on its own, and the combination of this data with physical measurements.

These user perception benchmarks are looked at in two ways, using an absolute method that takes a standard value and then compares a building's scores to it, and a relative method that relates the building scores to a sample of the New Zealand building stock. The results from Buildings 21 and 29 from the pilot database are used to demonstrate the practicality of these options. Building 21 is a low scoring building and Building 29 a high scoring building. This allows for a test of the sensitivity of the benchmarking method.

As the results from these buildings are being looked at in further detail, the individual distributions are included in Appendix G, so that reference can be made to the data. The key features of the buildings and the survey results are displayed below in Table 36.

Table 36: Building features

	Building 21	Building 29
Location	Wellington	Wellington
Building Use	Commercial	Commercial
Building Age	1970-2000	>2000
Building Height	3-6 floors	>6 floors
BEES floor area grouping	4 (3,500m ² - 9,000m ²)	4 (3,500m ² - 9,000m ²)
Ventilation	AC	MM
Design Intent	Conventional	Sustainable
Number of Surveys	32	192
Response Rate	56%	80%
Who carried out survey	E ³ BW	Author
Year Surveyed	2006	2010

Comparing the two buildings there are many differences in terms of characteristics and survey results. The low scoring building (Building 21) is older, and fully air conditioned with a lower response rate and overall lower number of respondents. Although they have similar floor areas, Building 29 has significantly more floors, which indicates that these floors are shallower, allowing the possibility of more natural ventilation and natural light over the floors, characteristics which occupants are thought to like.

6.1 Questionnaire variables

Generally, rating tools allocate a number of points to different aspects of buildings and combine them to come to an overall building score. The part of the current NZ rating tool method, Green Star NZ Office rating tool most applicable to the types of questions asked in the questionnaire would be the section on the indoor environmental quality, as mentioned in **Section 3.5: Current building rating tools**. The indoor environmental quality section deals with temperature, lighting, noise and in the equivalent performance tool could also look at the physical space the occupants are working in.

Due to the many variables that can affect the indoor environment, and the occupants' rating of it, an appropriate indicator set needs to be determined with which to assess the indoor environment. The different categories that make up the indoor environmental quality section in the rating tools briefly introduced in **Section 3.5: Current building rating tools** have corresponding questions in the questionnaire.

Therefore the summary variable questions that correspond to the indoor environmental quality that are considered in this assessment are:

- *Temperature Overall in Winter*
- *Temperature Overall in Summer*
- *Air Overall in Winter*
- *Air Overall in Summer*
- *Lighting Overall*
- *Noise Overall*
- *Overall Comfort*
- *Space*

A study that focussed on the quantification of occupant comfort and the practicality of combined indices determined that:

'Ranking several buildings in order of merit, by using a combination of several aspects evaluated by the occupants, seems unlikely to rank them in the order that would have been obtained by asking them directly about the buildings' overall merit.... It seems prudent, then, to continue to consider each aspect separately against norms that have been developed rather than rely solely on an overall evaluation' (Humphreys, 2005, p. 325).

It was thus determined that the individual variables as well as *Overall Comfort* should be included.

6.2 Building 21 and Building 29 ratings with BUS benchmarks

As stated earlier, Building 29 is a high scoring building, while Building 21 is low scoring. To fully understand the potential of the benchmarks that will be explored subsequently, the scores are presented against the current BUS benchmarking method, as outlined in **Section 3.4 Assessment of BUS benchmarks**. This includes the direct comparison of the individual scores to the NZ 2010 benchmark and the equivalent 'Better', 'Similar', or 'Worse' rating, and the overall BUS rating score out of 7 for both the 'all variable' and 'selected variable' methods.

Table 37: Current benchmark scores

Variable	Building 21	Building 29
Temperature Overall in Winter	Similar	Better
Temperature Overall in Summer	Similar	Better
Air Overall in Winter	Worse	Better
Air Overall in Summer	Similar	Better
Lighting Overall	Similar	Better
Noise Overall	Worse	Better
Overall Comfort	Worse	Better
Space	Similar	Better
BUS rating score: selected variable method	3	7
BUS rating score: all variable method	3	6

From these methods, Building 29 scores the highest possible ratings for all but the ‘All variable rating score method.’ Building 21, although lower scoring, is not scoring the lowest possible. These results will be able to be compared to later when exploring potential benchmarks.

6.3 Absolute benchmarks

Due to the nature of the scale, and the restriction in the possible ratings (1-7), the resultant scores can be explored solely in relation to the scale, without reference to the performance of the rest of the NZ building stock. These are referred to as absolute benchmarks in this report, and are largely methods of summarising the data into a comparable term.

6.3.1 Percentage of people satisfied/dissatisfied

As briefly described in **Section 3.3: Benchmarks of people’s perceptions** and **Section 3.5: Current building rating tools**, both of which analyse how different questionnaire based methods benchmark their results, it was found that a large number focus on the ‘percentage of people satisfied or dissatisfied’ as a comparable term. For example, the questionnaire based methods used by AMA, TOBUS, CBE, and HOPE. This approach is the sole method that is used in the current BSRTs that incorporated people’s perceptions. The data received and analysed for Buildings 21 and 29 was split into two groups for each variable, people who were satisfied and people who were dissatisfied.

All of the questions have an ideal score of seven. Occupants who scored one, two, or three were considered dissatisfied, while those who rated four and above were considered satisfied with that particular variable. The responses were split in this way due to the nature of the scale. As the scales are 7-point they range from scores of one as ‘unsatisfactory’ to scores of seven as ‘satisfactory’ with the ‘break-even’ or neutral point a score of four. Four was assigned to the satisfactory side as:

‘For many environmental features, a neutral rating is actually a good indicator and a desirable score. People generally perceive their physical work setting as a backdrop to their activities. It tends to be noticed only when it is negative or objectionable’ (Aronoff & Kaplan, 1995, p. 348).

The scores for the corresponding questions are shown below for the two buildings in Figure 71 and 72.

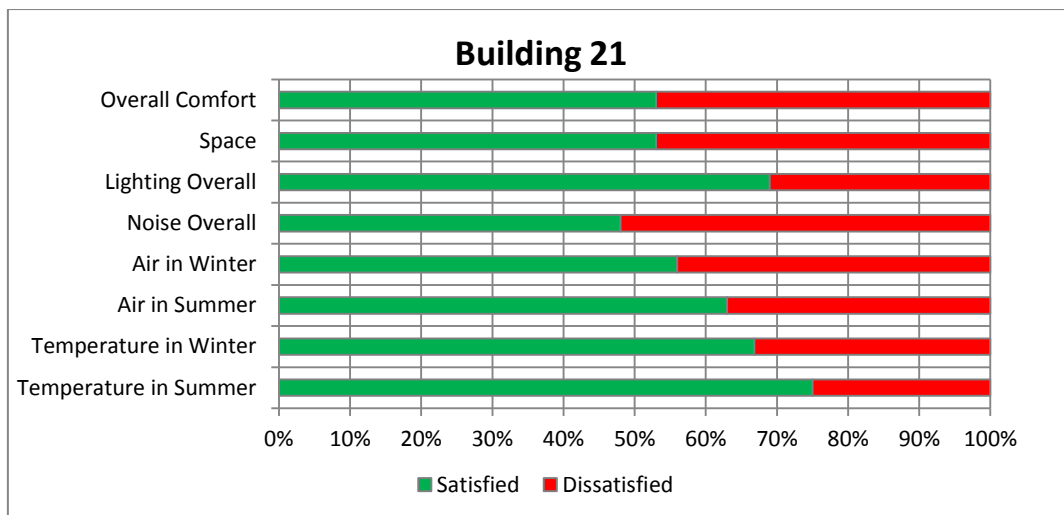


Figure 71: Percentage of people satisfied/dissatisfied - Building 21

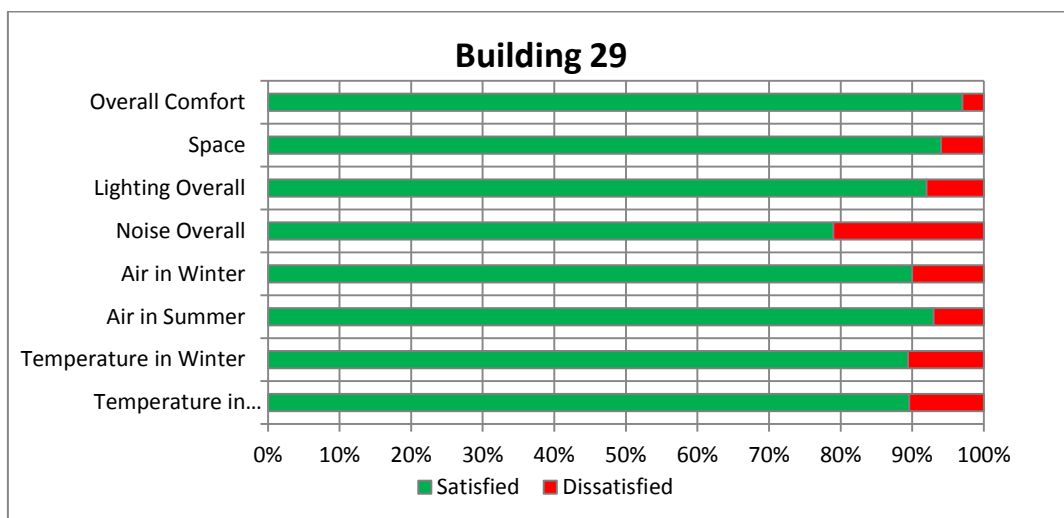


Figure 72: Percentage of people satisfied/dissatisfied - Building 29

In the comparison of the two examples, it can be seen that far more people are satisfied with Building 29 than Building 21. The average percentage for Building 29 is 90%, while only an average of 56% of occupants consider themselves to be satisfied in Building 21.

The results are skewed towards the positive because of including the midpoint as 'satisfactory' from there being more opportunity to rate the building as 'satisfactory' than 'unsatisfactory.'

Relating these results to the ASHRAE acceptable level of 80% satisfied (ASHRAE, 2004), for example, Building 29 complies for all but one variable, *Noise Overall* with only 79% satisfied. Building 21, however, is below this standard for every variable with the most satisfied variable being *Temperature Overall in Summer* with 75%.

There are several specific benefits to this method, in addition to the general benefits of absolute benchmarks. Using the percentage of people satisfied it can be directly linked and compared to the ASHRAE standards of 80% and the PMV and PPD tool. People are also easily able to understand what the number is referring to, as it is an accessible term to which everyone is able to relate. It is also related to the respondent's individual score rather than an average for the building.

The results could be combined in different ways. For example, *Temperature Overall in Winter* and *Summer* could be averaged for a thermal comfort score, or the questions could be split into two groups; the occupants' assessment of the different variables and then their assessment on how these variables affect them.

There is an argument for saying that this is the only method needed as it deals directly with the satisfaction of the individual occupant. Aronoff and Kaplan weigh in on this argument.

‘Diagnostic analyses, whether they involve tens or thousands of occupants, are more successful if they focus on assessing the workplace experienced by each individual rather than looking at average conditions suited to average people – who don’t exist’ (Aronoff & Kaplan, 1995, p. 353).

6.3.2 Individual Scores

When first invented, the Likert scale was used to assess attitudes of responses and to measure not only whether occupants like a building, but the way in which they like it. Although the questions from these are generally phrased in terms of agreement with a statement, which is different from the BUS questionnaire, the method in which the results are summarised can be applied to the data from the pilot database.

‘For each of the agree/disagree statements a number is attached to the response, and the scores from the positive are tallied, respectively. Degrees of “agreement” are possible with this technique. For example, if “strongly agree” scores 5, “agree” scores 4, neither/nor” scores 3, “disagree” scores 2 and “strongly disagree” 1, then if six people respond, and all strongly agree, then this statement scores 30 points’ (Roaf, 2004, p. 454).

All of the questions that are being focussed on in this section are numbered from one to seven, with seven being the ideal score. The individual ratings of the occupants are added up to create a total score, and this is able to form a percentage score by comparing the total score to the total scores possible.

Table 38: Individual scores method

Building 21

Variable	N	Score	Max Possible (7*N)	Percentage
Temperature in Summer	24	99	168	58.9%
Temperature in Winter	30	116	210	55.2%
Air in Summer	27	107	189	56.6%
Air in Winter	32	115	224	51.3%
Lighting	32	135	224	60.3%
Noise	31	107	217	49.3%
Overall Comfort	32	119	224	53.1%
Space	32	125	224	55.8%
Total		923	1680	54.9%

Building 29

Variable	N	Score	Max Possible (7*N)	Percentage
Temperature in Summer	144	763	1008	75.7%
Temperature in Winter	188	1032	1316	78.4%
Air in Summer	142	764	994	76.9%
Air in Winter	190	1032	1330	77.6%
Lighting	189	1071	1323	81.0%
Noise	189	906	1323	68.5%
Overall Comfort	190	1091	1330	82.0%
Space	191	1088	1337	81.4%
Total		7747	9961	77.8%

This results in a similar score for Building 21, compared to the average of 56% satisfied. The score for Building 29 is somewhat less than the average 90% satisfied as seen in **Section 6.3.1: Percentage of people satisfied/dissatisfied**. This 'individual scores' method, however, would to some degree be affected by the number of respondents.

In the 'Percentage of people satisfied' method the same weighting was applied to a neutral score of four and an ideal score of seven. This 'Individual score' method takes into account the differences between the scales, and the fact that the respondent made a conscious effort to rate a variable with a particular value. It does, however, place an almost unreachable target of 100%, as every occupant would have to rate a variable with a score of seven, which is highly unlikely.

6.3.3 Scores on the 7-point scale

All of the questionnaire based methods studied used the mean as the measure of central tendency, and to summarise the data received. For that reason the mean scores from Buildings 21 and 29 were examined further.

In terms of absolute benchmarks, one way the mean of the questionnaire data received can be explored is in regards to where the mean score lies on the spectrum of the scale. The HOPE project uses divisions in their scales to summarise the results from an occupant questionnaire to an overall building rating (Roulet, Flourentzou, Foradini, Bluysen, Cox, & Aizlewood, 2006).

Each of the indoor environmental quality questions in the BUS questionnaire focussed on, have an ideal score of seven so the same divisions can be made in the scale for each question. The scales were split into the different ranges and points were assigned to different bands.

The HOPE project when combining different performance areas in a building split the seven point scale into four different category ratings, discussed in **Section 3.5: Current building rating tools**. This was applied to the data received from Buildings 21 and 29. Further to this splitting into bands of quarters, fifths, and sixths was also explored.

Roaf considers the effect of different categories of a scale, and emphasises its importance:

'The perception of the relative performance of a building for a particular indicator can be packaged very differently according to the scale used on its yardstick. If the yardstick is divided into three categories: good (66-100 per cent), medium (33-66 per cent) or bad (0-33 per cent), then a building can be labelled as good that is actually only 67 per cent good on a percentage scale and would register as OK on a seven point scale. So the grain of the scale (course, medium or fine) can have a significant impact on the perceived performance of the product against an indicator yardstick' (Roaf, 2004, p. 46).

In **Section 3.3: Benchmarks of people's perceptions**, it was shown that several questionnaire based methods include the quartiles of the databases while the BUS method uses quintiles. It was determined through these studies that four or five groups were adequate, so these were adopted and analysed further. The scale was then also split into six as it reflects the nature of the 7-point scale (i.e. six divisions between the seven points). These are shown below in Table 39.

For the purposes of this illustration an arbitrary number of points were assigned to each band, with a minimum of 1 in all cases.

Table 39: Bands of the scale

HOPE	No. of Points	Quarters	No. of Points	Fifths	No. of Points	Sixths	No. of Points
1.00-1.99	1	1-2.50	1	1-2.20	1	1.00-1.99	1
2.00-4.00	2	2.51-4.00	2	2.21-3.40	2	2.00-2.99	2
4.01-5.50	3	4.01-5.50	3	3.41-4.60	3	3.00-3.99	3
5.51-7.00	4	5.51-7.00	4	4.61-5.80	4	4.00-4.99	4
				5.80-7.00	5	5.00-5.99	5
						6.00-7.00	6

Figures 73-76 illustrate the different splits in the scale and how the distribution of scores from the pilot database is spread over the scale, using the question for *Overall Comfort* as an example.

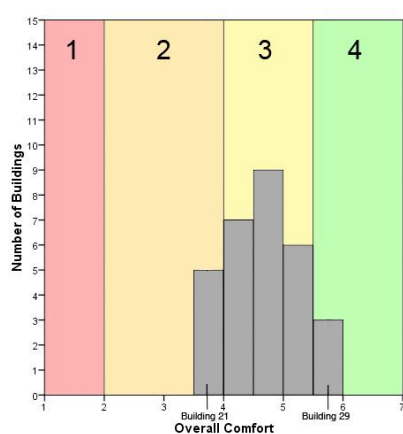


Figure 73: HOPE

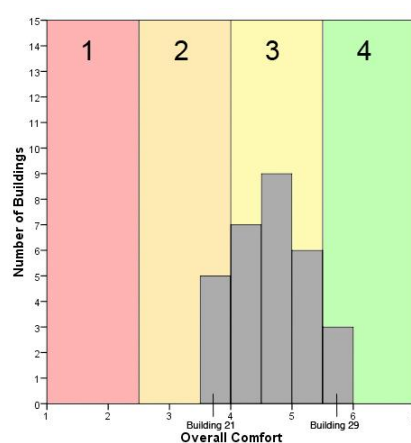


Figure 74: Quarters

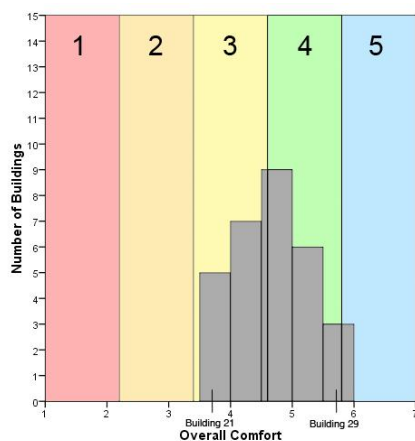


Figure 75: Fifths

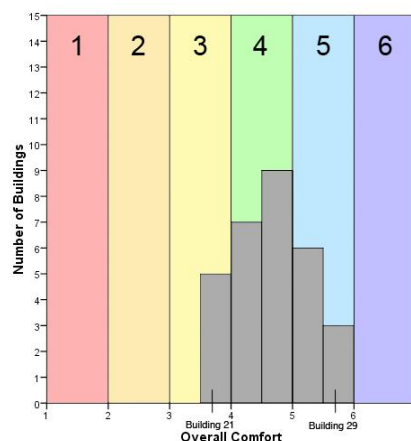


Figure 76: Sixths

The mean scores for Buildings 21 and 29 of the questions were then compared to these bands to determine where they rate, and how many points they receive, using the assigned number of points described above.

Table 40: Benchmarks from absolute scale method

Building 21

Variable	Score	HOPE	Quarters	Fifths	Sixths
Temperature in Summer	4.12	3	3	3	4
Temperature in Winter	3.87	2	2	3	3
Air in Summer	3.96	2	2	3	3
Air in Winter	3.59	2	2	3	3
Lighting	4.22	3	3	3	4
Noise	3.45	2	2	3	3
Overall Comfort	3.72	2	2	3	3
Space	3.91	2	2	3	3
Total Points Achieved		18	18	24	26
Maximum Possible		32	32	40	48
Percentage		56.3%	56.3%	60.0%	54.2%

Building 29

Variable	Score	HOPE	Quarters	Fifths	Sixths
Temperature in Summer	5.32	3	3	4	5
Temperature in Winter	5.52	4	4	4	5
Air in Summer	5.40	3	3	4	5
Air in Winter	5.46	3	3	4	5
Lighting	5.70	4	4	4	5
Noise	4.81	3	3	4	4
Overall Comfort	5.75	4	4	4	5
Space	5.73	4	4	4	5
Total Points Achieved		28	28	32	39
Maximum Possible		32	32	40	48
Percentage		87.5%	87.5%	80.0%	81.3%

The total number of points received by the buildings was then compared to the total number of points possible and the percentage found. The results for Building 29 from this method rates similarly to the average of percentage satisfied in the first 'Percentage of people satisfied/dissatisfied' method.

This could then be incorporated, with or without physical measurements, into a rating tool similar to Green Star NZ. Weightings could also be applied to each variable to place more importance on some than others, as at present they each have equal weighting.

The difference between the divisions is quite small, in terms of overall scoring, with the largest difference resulting in a difference of 7.5%.

In terms of which divisions in the scale are most appropriate there is not a definitive answer, but some have more advantages than others. In comparison between the fourths and the fifths, the fourths seems to be slightly better as it arguably breaks the scale in more logical places. All of them also take into account the change from unsatisfied to satisfied at the 'break-even' midpoint of the scale which is essential in the nature of the scale, except for the division into fifths. Referring back to **Section 5.3: Overview of scores of the buildings**, it can be seen that in the pilot database the large majority of the mean building scores are within the range of 4-5.5, and there are not many buildings scoring below the midpoint. This could then be used as a different division which reflects the nature of the scale and with regard to the distribution of the scores. The splitting into sixths reflects the bands of the original scale, with 7-points and six bands.

Building 29 is at the top of the distribution for a large number of the questions and scores highly in the different bands. Building 21 is at the bottom of the distribution and the lowest it scores is two.

The way the numbers are assigned, a building can still accumulate seven points when scoring very badly in all variables. This could be potentially moved down by assigning no points if the building mean score lies in the bottom end of the scale.

6.3.4 Summary of absolute benchmarks

The main advantage of absolute benchmarks is the fact that a building can be assessed independently without the need for a representative database which can be difficult to obtain. However, if the benchmark is determined solely in regard to the points on the scale, what needs to be ensured is that it is possible to achieve the highest and lowest bands. Using the pilot database and the questions that make up the indoor environment, for all of the splits apart from the sixths there are some buildings that are scoring in the highest band, but there are none in the bottom over all groupings. A more representative sample, inclusive of some lower scoring buildings will need to be obtained first to determine if this approach is feasible.

Oseland (2007) also discusses a disadvantage of absolute benchmarks:

'If the survey is post project only, it may not be clear whether the feedback indicates improvement or not, because the starting point is unknown. For example, finding 85% satisfaction with a particular aspect of the building may seem high, but there is no guarantee it was not higher before. If there is no pre-project survey data, the alternative is to benchmark the result with those found in the rest of the portfolio, or with the buildings of other similar organisations' (Oseland, 2007, p. 38).

These absolute benchmarks could be, and often are, combined to make a database of buildings rated with the same benchmark method. This reduces the negative aspect of the absolute benchmarks that they are isolated from the actual buildings stock. It then becomes a discussion of whether to compare the scores from a building at the individual question level straight away, or summarise the buildings in terms of the scale and how they are rating, and then compare them to a database of representative buildings.

For these particular absolute benchmarks explored, the type of scale they employ is paramount. The percentage of people satisfied can only be determined from scales that have a break-even score, which are the questions that have an ideal score of one or seven. Those with an ideal score of four would not be compatible due to the fact that the scale is bipolar. Using the mean score and dividing the scales could potentially be used for the questions with an ideal score of four, however, their distributions would be different and the divisions would need to be changed, resulting in them remaining not directly compatible with the unipolar scales.

Absolute benchmarks would also have the additional benefit of not having to be constantly updated as more buildings were added to the database.

The different splits in the bands do not have a large effect on the overall score, although this could be due to the distributions of the scores not being very wide ranging.

6.4 Relative Benchmarks

Another way of approaching benchmarking of user perceptions is to relate the scores to the ratings of similar buildings in New Zealand as a first step. The following section suggests possible ways of doing so.

6.4.1 Distribution of scores in relation to the pilot database

The mean scores from Building 21 and Building 29 were compared to the quartiles and quintiles of the mean building scores of the pilot database.

The idea of creating performance bands was first referred to in **Section 3.2.1: Energy benchmarks**. In this example a cumulative distribution was created which allowed the percentiles to be an indicator of the energy performance of the building.

The performance bands were emulated in the benchmarking of occupants' perceptions. The OPN methodology results in a radial diagram that includes the quartiles of the database so the results of the individual building can be compared to the larger dataset. AMA WorkWare also uses quartiles as a point of comparison, stating with their results in which quartile of a larger dataset an individual fits. BUS do something similar, but use quintiles as opposed to quartiles, as well as stating the actual relevant percentile the building achieves.

Oseland (2007) refers to different ways of benchmarking results from Post Occupancy Evaluations. He states:

'Another useful way of benchmarking the questionnaire results is to order all the buildings in the database according to their overall rating of satisfaction or other index. Where the building being assessed is ranked in the database is then highlighted, or at minimum it is identified whether it is ranked in the upper, lower or mid-quartile range' (Oseland, 2007, p. 39).

Oseland (2007) also offers quintiles as an option in terms of bands of performance:

'Use a BREEAM-style rating where, for example, the overall ratings could be classified as 'very poor' (<20%), 'poor' (20-40%), 'okay' (40-60%), 'good' (60-80%), and 'excellent' (>80%). The percentages in parentheses could represent either the percentage satisfied or the score on another index, the percentile where the overall rating lies in the database, or the ranked order of the building' (Oseland, 2007, p. 39).

Figures 77 and 78 illustrate where the different bands split the pilot database in terms of *Overall Comfort*. The same numbers were applied to the different bands as before. As noted previously these are a matter of choice and could be selected depending on preference, but these selected here do indicate clearly which band the buildings' mean scores fit into.

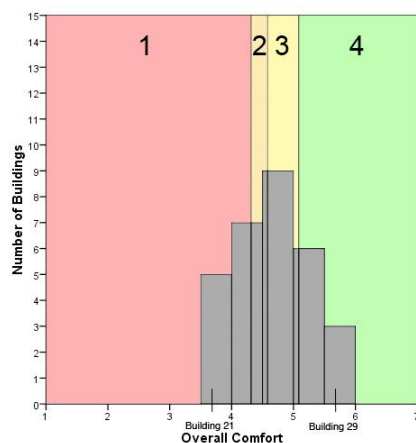


Figure 77: Quartiles

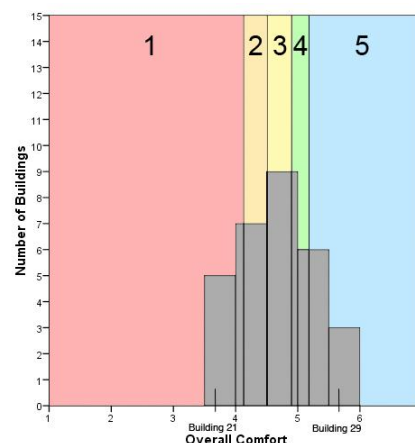


Figure 78: Quintiles

Table 41: Benchmarks for quartile/quintile method

Building 21

Variable	Score	Quartile	Quintile
Temperature in Summer	4.12	2	2
Temperature in Winter	3.87	1	1
Air in Summer	3.96	2	2
Air in Winter	3.59	1	1
Lighting	4.22	1	1
Noise	3.45	1	1
Overall Comfort	3.72	1	1
Space	3.91	1	1
Total Points Achieved		10	10
Maximum Possible		32	40
Percentage		31.3%	25.0%

Building 29

Variable	Score	Quartile	Quintile
Temperature in Summer	5.32	4	5
Temperature in Winter	5.52	4	5
Air in Summer	5.40	4	5
Air in Winter	5.46	4	5
Lighting	5.70	4	5
Noise	4.81	4	5
Overall Comfort	5.75	4	5
Space	5.73	4	5
Total Points Achieved		32	40
Maximum Possible		32	40
Percentage		100.0%	100.0%

For Building 29, the results of 100% of the total possible points able to be obtained, indicates that Building 29 is one of the highest performing buildings in the database. This method gave quite different results than the absolute benchmarks previously explored, where they received scores of about 80%. This ideal scoring is almost an overoptimistic rating of the building's performance. Although it rates in the highest percentile group for each question when looking at the mean score, there is room for improvement and the score of 100% does not represent this.

Building 21 gave quite different results, with the resulting percentage score around the 30% mark, scoring in the lowest percentile grouping the large majority of the time. In terms of the comparison to the absolute benchmark previously explored this is a significant amount lower (compared to about 50%).

6.5 Combined benchmarks

6.5.1 In relation to the scales and the pilot database

Due to the large differences between the relative and the absolute benchmarks, an attempt was made to incorporate both the nature of the scale and the distribution of the NZ buildings. The methods with the most similarities are the quarters and the quartiles, as well as the fifths and the quintiles. These were then averaged to obtain a combined score.

Table 42: Benchmarks for combined scale and percentile method

Building 21

Variable	Score	Quarters	Quartile	Total	Fifths	Quintile	Total
Temperature in Summer	4.12	3	2	5	3	2	5
Temperature in Winter	3.87	2	1	3	3	1	4
Air in Summer	3.96	2	2	4	3	2	5
Air in Winter	3.59	2	1	3	3	1	4
Lighting	4.22	3	1	4	3	1	4
Noise	3.45	2	1	3	3	1	4
Overall Comfort	3.72	2	1	3	3	1	4
Space	3.91	2	1	3	3	1	4
Total Points Achieved		18	10	28	24	10	34
Maximum Possible		32	32	64	40	40	80
Percentage		56.3%	31.3%	43.8%	60.0%	25.0%	42.5%

Building 29

Variable	Score	Quarters	Quartile	Total	Fifths	Quintile	Total
Temperature in Summer	5.32	3	4	7	4	5	9
Temperature in Winter	5.52	4	4	8	4	5	9
Air in Summer	5.40	3	4	7	4	5	9
Air in Winter	5.46	3	4	7	4	5	9
Lighting	5.70	4	4	8	4	5	9
Noise	4.81	3	4	7	4	5	9
Overall Comfort	5.75	4	4	8	4	5	9
Space	5.73	4	4	8	4	5	9
Total Points Achieved		28	32	60	32	40	72
Maximum Possible		32	32	64	40	40	80
Percentage		87.5%	100.0%	93.8%	80.0%	100.0%	90.0%

The resultant scores were remarkably similar to each other, in regards to the number of splits. There was a 2.5% difference between the total of the quartile and quintile method for both buildings.

This method gives a logical answer for Building 29 as it takes into account that the building is very high performing in terms of other New Zealand buildings, but it still indicates its future potential in terms of mean score on the scale. The downside of this method is that it is again dependent on the assigning of values to the different bands. The worst performing buildings could still get an overall number of points of 14, even if it was in the bottom number of buildings and lowest scale rating. This could again, be reduced by changing the minimum number of points.

6.5.2 Alternate way of incorporating both the scale and the benchmark

In an attempt to incorporate both the scale qualities and the database, the already established BUS rating scores method, introduced in **Section 3.4: Assessment of BUS benchmarks**, was re-examined. Currently, the method looks at the total score over all variables such as design, comfort, space etc. Instead of using this one score over every variable it was rearranged to make it more compatible with the IEQ section of the Green Star NZ Office tool, and the corresponding questions in the questionnaire.

As discussed in **Section 3.4: Assessment of BUS benchmarks** there are two methods, one that focuses on ten summary variables and another that includes all questions in the questionnaire.

The BUS equivalent score relates to the BUS bands identified earlier, which are:

- 1 (0-14.3%) Very poor
- 2 (14.4-28.6%) Poor
- 3 (28.7-42.9%) Below average
- 4 (43-57.2%) Average
- 5 (57.3-71.4%) Above average
- 6 (71.5-85.7%) Good practice
- 7 (85.8-100%) Exceptional

For Building 21, displayed below in Table 43, the overall rating of the building using the 'BUS rating score: selected variable method' results in an overall score of 37.5%.

Table 43: Building 21 - BUS rating score: selected variable method

	Rating Score	
Variables	Temperature in Summer	3
	Temperature in Winter	2
	Air in Summer	2
	Air in Winter	1
	Lighting	3
	Noise	1
	Overall Comfort	1
	Space	2
Summary Info	Total Points Achieved	15
	Maximum Possible (5*8)	40
	Percentage (Total/40)	37.5%
	BUS equivalent rating score	3

For Building 29, the eight variables that make up the BUS rating score: selected variables method, all, similar to the quartile/quintile method, get a rating score of five and hence an overall score of 100%.

The advantage of using the BUS rating score method is that it has two options, one of which allows for the inclusion of the sub-questions, those dealing with the source and the type of problems the occupants are facing. The relevant questions were then analysed further using the 'BUS rating score: all variable method', and an individual total score was calculated for each variable. These are displayed in Table 44 and 45 below.

Table 44: Building 21 – BUS rating score: all variable method

	Rating Score		Rating Score	
Variables	TW overall	2	TS overall	3
	TW hot/cold	1	TS hot/cold	2
	TW stable/varies	1	TS stable/varies	3
Summary Info	Total Thermal Winter (15)	4	Total Thermal Summer(15)	8
	Maximum Possible (5*3)	15	Maximum Possible (5*3)	15
	Percentage (Total/15)	26.7%	Percentage (Total/15)	53.3%
	BUS equivalent rating score	2	BUS equivalent rating score	4
Variables	AW still/draughty	5	AS still/draughty	3
	AW dry/humid	3	AS dry/humid	3
	AW fresh/stuffy	1	AS fresh/stuffy	1
	AW odourless/stuffy	3	AS odourless/stuffy	3
	AW overall	1	AS overall	2
Summary Info	Total Air Winter	13	Total Air Summer (25)	12
	Maximum Possible (5*5)	25	Maximum Possible (5*5)	25
	Percentage (Total/25)	52.0%	Percentage (Total/25)	48.0%
	BUS equivalent rating score	4	BUS equivalent rating score	4
Variables	Lighting Overall	3	Noise Overall	1
	Natural Light	1	Noise from Colleagues	1
	Glare from sun and sky	5	Noise from other people	2
	Artificial Light	1	Other noise from inside	2
	Glare from lights	4	Noise from outside	1
			Unwanted interruptions	3
Summary Info	Total Lighting (25)	14	Total Noise (30)	10
	Maximum Possible (5*5)	25	Maximum Possible (5*6)	30
	Percentage (Total/25)	56.0%	Percentage (Total/30)	33.3%
	BUS equivalent rating score	4	BUS equivalent rating score	4
Variables	Space	2	Total Thermal Winter	4
	Meeting Rooms	1	Total Thermal Summer	8
	Storage	1	Total Air Winter	13
	Furniture	3	Total Air Summer	12
	Space at desk	1	Total Lighting	14
			Total Noise	10
Summary Info	Total Office (25)	8	Total	69
	Maximum Possible (5*5)	25	Maximum Possible (5*32)	160
	Percentage (Total/25)	32.0%	Percentage (Total/160)	43.1%
	BUS equivalent rating score	3	BUS equivalent rating score	4

Table 45: Building 29 – BUS rating score: all variable method

	Rating Score		Rating Score	
Variables	TW overall	5	TS overall	5
	TW hot/cold	3	TS hot/cold	4
	TW stable/varies	5	TS stable/varies	5
Summary Info	Total Thermal Winter (15)	13	Total Thermal Summer(15)	14
	Maximum Possible (5*3)	15	Maximum Possible (5*3)	15
	Percentage (Total/15)	86.7%	Percentage (Total/15)	93.3%
	BUS equivalent rating score	7	BUS equivalent rating score	7
Variables	AW still/draughty	1	AS still/draughty	3
	AW dry/humid	3	AS dry/humid	4
	AW fresh/stuffy	5	AS fresh/stuffy	5
	AW odourless/stuffy	5	AS odourless/stuffy	5
	AW overall	5	AS overall	5
Summary Info	Total Air Winter	19	Total Air Summer (25)	22
	Maximum Possible (5*5)	25	Maximum Possible (5*5)	25
	Percentage (Total/25)	76.0%	Percentage (Total/25)	88.0%
	BUS equivalent rating score	6	BUS equivalent rating score	7
Variables	Lighting Overall	5	Noise Overall	5
	Natural Light	1	Noise from Colleagues	3
	Glare from sun and sky	4	Noise from other people	4
	Artificial Light	3	Other noise from inside	5
	Glare from lights	5	Noise from outside	1
			Unwanted interruptions	4
Summary Info	Total Lighting (25)	18	Total Noise (30)	22
	Maximum Possible (5*5)	25	Maximum Possible (5*6)	30
	Percentage (Total/25)	72.0%	Percentage (Total/30)	73.3%
	BUS equivalent rating score	6	BUS equivalent rating score	6
Variables	Space	5	Total Thermal Winter	13
	Meeting Rooms	5	Total Thermal Summer	14
	Storage	5	Total Air Winter	19
	Furniture	5	Total Air Summer	22
	Space at desk	1	Total Lighting	18
			Total Noise	22
Summary Info	Total Office (25)	21	Total	129
	Maximum Possible (5*5)	25	Maximum Possible (5*32)	160
	Percentage (Total/25)	84.0%	Percentage (Total/160)	80.6%
	BUS equivalent rating score	6	BUS equivalent rating score	6

As each section is made up of different numbers of questions, the percentage of the total points received compared to the total possible was focussed on instead, as it is more comparable. The average percentage rating over the different variables is for Building 21 is 43.1% and for Building 29 is 81.9%.

There are several advantages of this method. An important one is that it uses an already established method, but rearranges it to fit specific criteria, and the fact that it allows for the addition of all the components as well as the summary variables.

The main method in this is how the ratings for the individual questions are constructed. These could be moved around under different headings for more specific use. Similar to the last method, it takes into account both the database of buildings and where the score sits on the scale.

There is always a difficulty when combining questions with different formats. In this situation as the ratings come from slightly different question formats, it is easier to get a rating of five in those with an ideal score of seven than it is for those with an ideal score of four.

6.5.3 Comfort Index

Indices were used in several questionnaire based methods featured in **Section 3.3: Benchmarks of people's perceptions**.

'An index is a set of questionnaire items that combines multiple related responses into a single rating or score, e.g. by summing or averaging. Using an index derived from several questions covering the same topic is considered a more reliable measure than using a single question on that topic' (Oseland, 2007, p. 33).

An already established method using this questionnaire data is the *BUS Comfort Index*. As discussed in **Section 3.4: Assessment of BUS benchmarks** the *Comfort Index* is made up of seven different variables. This index uses the benchmark scores in its construction so is dependent on these being available. In order to calculate the *Comfort Index* and make it compatible with the other methods discussed it needs to be calculated from this database, using the data from Building 21 and 29.

Building 21 *Comfort Index* using pilot database:

Equation 9: *Comfort Index* for Building 21 using pilot database

$$\text{Comfort Index} = -1.38 = \frac{\text{Average of z-scores of ...}}{7}$$

$$= \frac{\frac{\text{ComfOver} (3.72 - 4.70)}{0.59} + \frac{\text{LtOver} (4.22 - 5.01)}{0.34} + \frac{\text{NseOver} (3.45 - 4.35)}{0.49} + \frac{\text{TSoVer} (4.12 - 4.29)}{0.49} + \frac{\text{TWOVer} (3.87 - 4.50)}{0.45} + \frac{\text{AirSOVer} (3.96 - 4.25)}{0.52} + \frac{\text{AirWOVer} (3.59 - 4.39)}{0.54}}{7}$$

Building 29 *Comfort Index* using pilot database:

Equation 10: *Comfort Index* for Building 29 using pilot database

$$\text{Comfort Index} = 1.90 = \frac{\text{Average of z-scores of ...}}{7}$$

$$= \frac{\frac{\text{ComfOver} (5.75 - 4.70)}{0.59} + \frac{\text{LtOver} (5.70 - 5.01)}{0.34} + \frac{\text{NseOver} (4.81 - 4.35)}{0.49} + \frac{\text{TSoVer} (5.32 - 4.29)}{0.49} + \frac{\text{TWOVer} (5.52 - 4.50)}{0.45} + \frac{\text{AirSOVer} (5.40 - 4.25)}{0.52} + \frac{\text{AirWOVer} (5.46 - 4.39)}{0.54}}{7}$$

The resultant indices are reflective of the general scoring of the building results so far; Building 21 scoring low, in the negatives, Building 29 scoring high, in the positives. The distribution of *Comfort Indices* was explored in an earlier section, **Section 5.4.2: BUS summary method comparison**, using both the UK and the NZ BUS databases. The impact that the benchmark database has on the *Comfort Index* is illustrated through the

comparison between the UK and NZ distributions. The results from these 30 buildings using the pilot database showed a range from -1.38 to +1.90, with a mean score of -0.002, Building 21 being the lowest, Building 29 being the highest. The distribution was split into quartiles and quintiles similarly to the individual questions.

Table 46: Splits in *Comfort Index* scale

Number of Points	Quartiles	Quintiles
1	-3.00 - -0.70	-3.00 - -0.73
2	-0.69 - -0.06	-0.72 - -0.40
3	-0.05 - 0.46	-0.39 - 0.26
4	0.47 - 3.00	0.27 - 0.59
5		0.60 - 3.00

Figures 79 and 80 illustrate the split in the data of the *Comfort Index*.

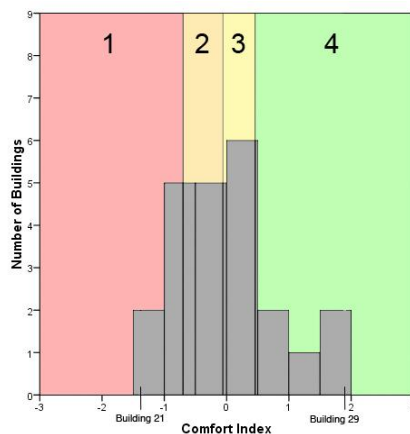


Figure 79: *Comfort Index* quartiles

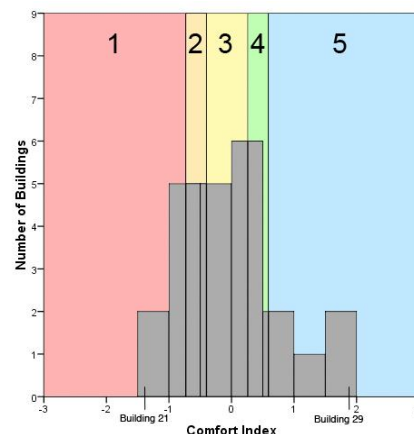


Figure 80: *Comfort Index* quintiles

With its *Comfort Index* score of +1.90, Building 29 sits in the top quartile and quintile of the dataset, so would receive the maximum number of points.

The benefit of this method is that the *Comfort Index* is an already established BUS method, created using the results from this BUS questionnaire. In terms of the distribution of the bands, the highest and lowest bands cover a large portion of the scale, while those in the middle are much narrower. This is largely to do with the general distribution of the scores and that the large majority of them are scoring around the midpoint of the scale. The bands could be changed accordingly to make them closer in terms of width.

6.5.4 Summary of relative and combined benchmarks

Relating the individual questions directly to the building stock, allows an analysis to be made at the question level as to how the individual variables are performing in regards to other buildings in New Zealand. This allows further opportunity for the problematic areas of the building to be determined, and the ability to state to a client where a particular variable of a building was performing in relation to other buildings in New Zealand.

Analysing the buildings in this way means a representative sample is needed from the outset. A profile of the existing building stock in New Zealand in terms of user perceptions should be the first step in creating benchmarks in this way.

If a similar technique to BUS is followed, the buildings that were used to form the benchmark would change as more buildings were added to the database. This would result in the benchmarks constantly changing; it would, however, reflect the changing nature of the occupants' ratings and their perceptions of the buildings that they work in.

6.6 User perception benchmarks summary

The average scores of the different benchmarking methods are displayed in Figure 81. This illustrates the differences between the three methods. The absolute benchmarks for the two buildings score much closer to each other than the relative and the combined benchmarks. The buildings that are rated solely in regard to where they fit in the distribution, the relative benchmarks, result in a lower score for Building 21 and a higher score for 29.

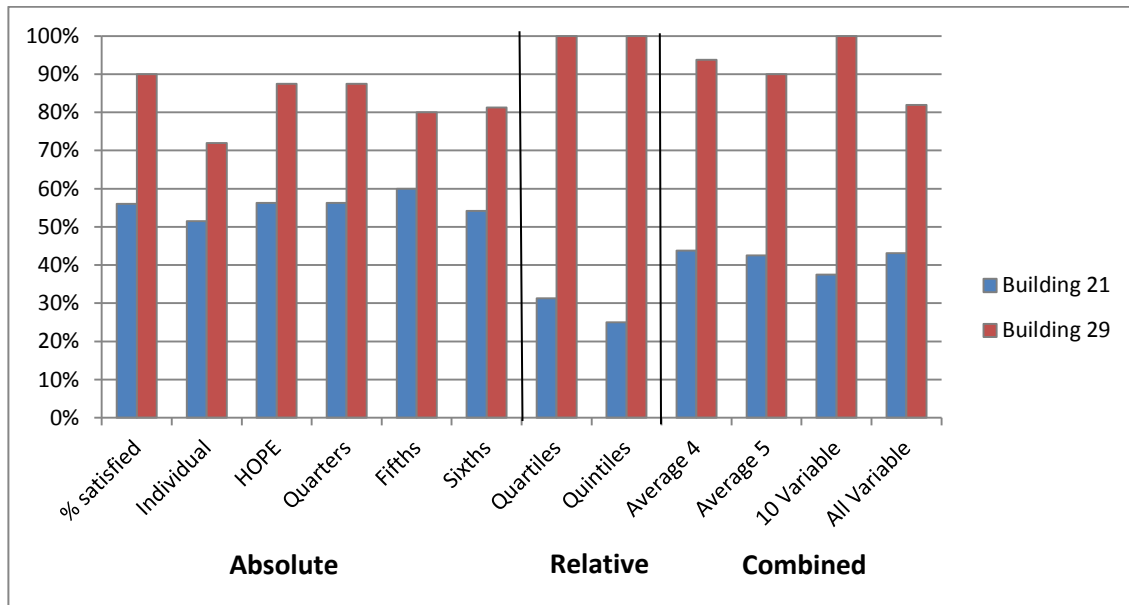


Figure 81: Comparison of different benchmarks for Absolute, Relative and Combined methods

These methods are by no means exhaustive, but are an exploration of the feasibility of some methods that could be used. The benchmarks trialled looked at each individual respondent's scores, the mean score in relation to the scale, the mean score in relation to the population, the mean of the last two methods combined (scale and population), and others based on already established methods. While the majority of the methods use the same indicator set, the variable rating score methods are slightly different so should be kept in mind in the comparisons.

The benchmarks explored for Building 21, ranged from 25% to 60%, and for Building 29 from 72% to 100%. There are large differences in the results from the way the scores are summarised and benchmarks produced. This reinforces the importance of standardised methods being used.

It seems from the data, that if the results are based solely on the building stock, with no regard to the qualities of the scale, that an overestimation of the performance could result. This suggests that any benchmarks that are created should not be based entirely on the percentiles of the pilot database, or should at least be checked again against a full representative sample.

From the limited sample of buildings in the pilot database it is difficult to make conclusive statements about the best method. In terms of which methods are the most promising, they each have their own advantages and disadvantages. The percentage satisfied method, however, which is the method currently incorporated in rating tools, would be a suggested method to explore further for New Zealand as it relates directly to the occupants, has minimal reliance on the influence of the researcher, and has an attainable 100% point.

6.7 Incorporation into rating tools

Once the benchmarks have been established the next issue to be considered is the feasibility of incorporating them into the relevant building rating tool, which in the case of New Zealand is NZ Green Star. The new performance Green Star tool that is in development is based on the NABERS methodology and approach of combining questionnaire data into a whole building performance rating tool. This was looked at as an example and one way of incorporating the data. There will of course be numerous other methods, but NABERS was deemed the most relevant to explore further. The development of benchmarks explored above has given an equal weighting to each variable that relates to the indoor environmental quality. NABERS gives different weighting for each variable, giving three times more importance on Thermal Comfort than Lighting Quality for example. NABERS includes five different variables in the assessment. These five variables and the percentage satisfied were analysed, closely following the style of the existing tool.

Each of the five variables has been given a maximum of 30 points. Four of these; Thermal Comfort, Air Quality, Acoustic Comfort and Lighting Quality, are a combination of physical measurements and questionnaire results, with the fifth Office Layout being made up of 30 points solely from the questionnaire. The percentage satisfied is then calculated for Thermal Comfort, Lighting, Noise, and Air Quality based on occupant responses to the 'overall' question in each of these categories using the procedure outlined in **Section 6.3.1: Percentage of people satisfied/dissatisfied** (Figures 71 and 72). In the case of Thermal Comfort and Air Quality, the percentage satisfied is derived by averaging the percentages for winter and summer. The resultant percentage satisfied is then multiplied by the number of points (15) to obtain a score for the questionnaire section. This and the number of points obtained from the physical measurements are added together to get a total rating out of 30 for each variable, and then the weighting for the different sections is applied.

As no physical measurements were taken in the buildings at the time of the survey, an arbitrary number of points (10 points out of 15) have been included to enable the demonstration to be carried out. The process and values used are displayed for Building 21 and 29 below in Table 47.

Table 47: Incorporation example using percentage satisfied

Building 21

Variable	Whole building	% satisfied	S points	token P points	S+P	Weightings	(S+P)* weighting	Total possible
Thermal Comfort	P(15pts) + S(15pts)	71.0%	10.65	10	20.65	30%	6.20	9
Air Quality	P(15pts) + S(15pts)	59.5%	8.93	10	18.93	20%	3.79	6
Acoustic Comfort	P(15pts) + S(15pts)	48.0%	7.20	10	17.20	20%	3.44	6
Lighting Quality	P(15pts) + S(15pts)	69.0%	10.35	10	20.35	10%	2.04	3
Office Layout	S (30pts)	53.0%	15.90	0	15.90	20%	3.18	6
Total							18.64	30
							62.1%	100%

Building 29

Variable	Whole building	% satisfied	S points	token P points	S+P	Weightings	(S+P)* weighting	Total possible
Thermal Comfort	P(15pts) + S(15pts)	89.5%	13.43	10	23.43	30%	7.03	9
Air Quality	P(15pts) + S(15pts)	91.5%	13.73	10	23.73	20%	4.75	6
Acoustic Comfort	P(15pts) + S(15pts)	79.0%	11.85	10	21.85	20%	4.37	6
Lighting Quality	P(15pts) + S(15pts)	92.0%	13.80	10	23.80	10%	2.38	3
Office Layout	S (30pts)	94.0%	28.20	0	28.20	20%	5.64	6
Total							24.16	30
							80.5%	100%

NB: S points: calculated by multiplying the percentage satisfied by the total number of S points e.g. 15pts or 30pts

NABERS then assigns an overall star rating for the indoor environmental quality of the building. The stars generally relate to:

Table 48: NABERS star ratings

1 star	Poor
2 stars	Below the median
3 stars	Above the median
4 stars	Excellent = top 20%
5 stars	Exceptional = roughly top 5%

There are general rules in the derivation of NABERS rating bands, as described in regards to Water Benchmarks by Bannister et al (2005):

'The NABERS rating uses bands that are based on the position of the building within the building population. There are a number of rules used to determine the rating bands for a particular population, being:

- 1. The rating scale should encompass at least 80% of the population*
- 2. The midpoint score should be based on the population median*
- 3. The full mark score should represent a level of efficiency essentially beyond normal technological solutions, but attainable through innovation, and*
- 4. The rating bands should be linear*

In the inevitable event of conflict between these requirements, the midpoint and linearity rules dominate the setting of the scale' (Bannister, Munzinger, & Bloomfield, 2005, p. 17).

From this, Building 21 received an overall score of 62.1%. Using the online rating tool 'NABERS office rating calculator', this gives a star rating of 3.5 stars. 3.5 stars is equivalent to:

'You are managing aspects of indoor environment, which reflects an awareness of the importance of indoor environment impacts on occupants. Improvements are still possible' (Australian Department of Environment, Climate Change and Water, 2010).

From this Building 29 received an overall score of 80.5%. Using the online rating tool 'NABERS office rating calculator', this gives a star rating of 4.5 stars. 4.5 stars is equivalent to:

'Your building demonstrates strong indoor environmental performance reflecting good equipment selection, operation and management' (Australian Department of Environment, Climate Change and Water, 2010).

This calculator is based on Australian buildings so relates to their needs and buildings rather than the building stock of New Zealand.

The other benchmarks explored could be incorporated in a similar way to NABERS, inputting the relevant figures in place of the input 'Percentage Satisfied.' Below is an example. It is a similar analysis for the altered BUS rating score: all variable method, using the overall percentage score (detailed on pages 126-127).

Table 49: Incorporation example using the altered BUS rating score: all variable method

Building 21

Variable	Whole Building	Percentage score	S points	token P points	S+P	Weightings	(\$+P)* weighting	Total possible
Thermal Winter	P(15pts) + S(15pts)	26.7%	4.01	10	14.01	15%	2.10	4.5
Thermal Summer	P(15pts) + S(15pts)	53.3%	8.00	10	18.00	15%	2.70	4.5
Air Quality Winter	P(15pts) + S(15pts)	52.0%	7.80	10	17.80	10%	1.78	3
Air Quality Summer	P(15pts) + S(15pts)	48.0%	7.20	10	17.20	10%	1.72	3
Lighting	P(15pts) + S(15pts)	56.0%	8.40	10	18.40	10%	1.84	3
Noise	P(15pts) + S(15pts)	33.3%	5.00	10	15.00	20%	3.00	6
Office Quality	S (30pts)	32.0%	9.60	0	9.60	20%	1.92	6
Total							15.06	30
							50.2%	100%

Building 29

Variable	Whole Building	Percentage score	S points	token P points	S+P	Weightings	(\$+P)* weighting	Total possible
Thermal Winter	P(15pts) + S(15pts)	86.7%	13.01	10	23.01	15%	3.45	4.5
Thermal Summer	P(15pts) + S(15pts)	93.3%	14.00	10	24.00	15%	3.60	4.5
Air Quality Winter	P(15pts) + S(15pts)	76.0%	11.40	10	21.40	10%	2.14	3
Air Quality Summer	P(15pts) + S(15pts)	88.0%	13.20	10	23.20	10%	2.32	3
Lighting	P(15pts) + S(15pts)	72.0%	10.80	10	20.80	10%	2.08	3
Noise	P(15pts) + S(15pts)	73.3%	11.00	10	21.00	20%	4.20	6
Office Quality	S (30pts)	84.0%	25.20	0	25.20	20%	5.04	6
Total							22.83	30
							76.1%	100%

This rating gives lower results for Building 29 than the Percentage Satisfied even though it is one of the higher performing buildings. This is most probably to do with the more detailed questions which are included. As discussed in **Section 3.4: Assessment of BUS benchmarks** earlier in the report, when all scores feature, the buildings generally do not rate as well. However, if the same process was carried out on all of the buildings in the pilot database it would still be one of the highest ratings.

Using the NABERS online rating tool, Building 21 with this different rating method received a score of 3 stars, which corresponds to:

'You are managing aspects of indoor environment, which reflects an awareness of the importance of indoor environment impacts on occupants. Improvements are still possible' (Australian Department of Environment, Climate Change and Water, 2010).

Using the online rating tool, Building 29 with this different rating method received a score of 4 stars, which corresponds to:

'Your building demonstrates strong indoor environmental performance reflecting good equipment selection, operation and management' (Australian Department of Environment, Climate Change and Water, 2010).

The weightings from the NABERS 'Whole Building' assessment were used in this example. Incidentally these weightings coincide with the strength of the correlations looked at in **Section 5.4: Questionnaire variables** between the summary variables and *Overall Comfort*. In the correlation *Temperature Overall in Winter* and *Temperature Overall in Summer* had the strongest correlation to *Overall Comfort* and have the highest weighting, followed by air quality. *Lighting Overall* had the weakest correlation with *Overall Comfort*.

7.0 CONCLUSIONS AND RECOMMENDATIONS

The overall aim of this project was to assist in the improvement of the performance of commercial and institutional buildings for their occupants through the exploration of user perception benchmarks. The research analysed the data received from a sample of surveyed New Zealand buildings to determine patterns in the data received and influencing variables within the questionnaire itself, and the buildings' design.

The objectives of this research were:

- To provide a general overview of existing studies carried out into user perceptions.
- To explore the nature and characteristics of benchmarks in general, those in which people's perceptions played a key part, and those already applied in the building industry.
- To create a pilot database of results from user questionnaires of New Zealand Buildings.
- To use the data collected to explore methods for determining benchmarks for NZ commercial and institutional buildings in terms of user perceptions.

The hypothesis was that it was feasible to develop user perception benchmarks for the establishment of country specific benchmarks. In terms of the data received there is plenty of opportunity for the data to be developed further and to be turned into benchmarks that are able to be incorporated into current building rating tools.

To investigate these aims and objectives the BUS occupant questionnaire was used to assess how the occupants perceived conditions in their buildings and any problems that they encountered, such as too much glare, or uncomfortable temperatures. A pilot database of results from the surveyed New Zealand buildings was compiled. This allowed flexibility in the statistical analysis, and exploration of how possible benchmarks could be constructed. These analyses enabled a discussion to take place on how the data could be made into benchmarks and how these could be incorporated into building rating tools.

There are many factors that can have a direct effect on the health, comfort and productivity of occupants. This is particularly important due to the long periods of time that occupants spend in buildings. There has been a vast amount of research carried out into the effect the indoor environment has on the occupants. This research project focussed on several main factors that make up the indoor environment; thermal conditions, air quality, lighting, and acoustic conditions.

An essential part of any improvement process is comparisons to standardised benchmarks or norms. This research is one of many steps for improving buildings for the ultimate end users of the buildings. The hope is that this will be able to be developed further with a larger, statistical reliable set of user perception benchmarks.

7.1 Conclusions

This section outlines the main findings and conclusions in relation to the specific aims and objectives of the research.

7.1.1 Determining users' perceptions

The first research objective was to provide a general overview of existing studies carried out into user perceptions.

Questionnaires have been proven to be an effective measure of assessing occupants' perceptions of the indoor environment. However, good practice needs to be followed in order for the questionnaires to be a viable measuring tool. The questionnaire based method needs to be standardised in order for benchmarks to be created from the data as small changes in the questionnaire can result in large differences in the results.

Using questionnaires as a tool for assessing the performance of a building is an effective way of seeing how it is affecting the occupants, and is cheaper and less time consuming than the technical measures that would be needed to reach the same level of detail. Questionnaires are able to be followed up with technical measurements if problems are determined. The USGBC, CIBSE, and ASHRAE recommend them as the first step in building performance assessment for assessing all existing buildings.

Numerous methodologies have been used to assess the indoor environment through the use of a questionnaire, several of which have become established over many years. Their use has confirmed that questionnaires are a valid and effective tool for obtaining an assessment of the indoor environmental quality. This brings the assessment to a more direct level than using technical measurements to assess the IEQ.

The literature review indicated there were several factors that need to be kept in mind in interpreting the results of assessing users' perceptions.

- The country that is being assessed will affect the results in terms of climate and culture.
- The way in which the occupants rate the buildings may change over time so the results cannot be compared to data that it is too old.
- Occupants tend to rate the overall questions with higher scores than the specific questions.
- Questionnaire design can have a direct effect on the responses received, so methods need to be standardised.
- Interpretation of responses needs to be standardised, using appropriate statistical methods.

7.1.2 Benchmarks

The second research objective was to explore the nature and characteristics of benchmarks in general, those in which people's perceptions played a key part, and those already applied in the building industry.

Benchmarks can take a variety of forms. They are points of comparison, and are a vital part in the improvement process, whether it is the improved performance of a building, or the improved running of a business.

Benchmarks should be simple enough to be easily used, but also need to offer enough information for an assessment of the building to be made. In terms of the indoor environmental quality, due to the many variables associated with it, key indicators need to be clearly determined. This project has focussed on *Temperature Overall in Winter, Temperature Overall in Summer, Air Overall in Winter, Air Overall in Summer, Lighting Overall, Noise Overall, Overall Comfort, and Space*, as representative of the internal environment.

Benchmarks are usually related to a larger dataset, or potential aspirational ratings. There have been many examples of the use of benchmarks through the building industry, and in the assessment of people's perceptions. They are generally based on simple statistics – a measure of central tendency, a measure of variability, and a measure of relative standing. A popular statistic used in benchmarks is percentiles as they allow for a specific value to fit in a larger dataset.

There are several benchmark methods used for assessing user perceptions related to buildings. The most highly developed of these are the Building Use Studies and Center for the Built Environment, Berkeley methods. These generally offer a profile of the building by displaying the scores from a range of variables, in a variety of ways.

The benchmarking process generally consists of assessing the building, choosing the right indicators on which to create the comparison, and then developing the benchmarks and targets against which the indicators can be compared.

7.1.3 Pilot database

The third objective of this research was to create a pilot database of results from user questionnaires of New Zealand Buildings.

The questionnaire data received from 30 buildings was compiled to form a pilot database. With a confidence level of 95% this database resulted in a margin of error of $\pm 18\%$. For a full statistically representative sample, 363 buildings around NZ would need to be surveyed.

The pilot database that was used for this study was compiled predominantly of buildings with large floor areas, and mainly situated in Wellington. This resulted in it not being representative of the general building stock, but still giving comparable results to other established methods. Using the mean results from these questionnaires creates a database, which is normally distributed for the different variables. This allows for parametric statistics to be applied. It also means that the methods used for determining benchmarks in general, in areas of building, and for people's perceptions are applicable to the data that is received from the BUS questionnaire.

The results from the occupant questionnaire data from this database followed consistent patterns to those of other established methods, with *Storage*, *Health* and the *Air* variables consistently scoring low. The highest scoring variables are for *Furniture*, *Image*, and *Lighting Overall*. *Image* is the variable with the largest variability, reflecting its subjective nature.

There are strong relationships between the vast majorities of the summary variables. Correlations were carried out on the summary variables, with particular focus on *Health*, *Overall Comfort*, and *Perceived Productivity*. These were highly correlated with each other and many of the other variables. In terms of the indoor environment variables, *Air Overall in Summer*, *Air Overall in Winter*, *Temperature Overall in Summer* and *Temperature Overall in Winter* were consistently the most highly correlated with *Health*, *Overall Comfort* and *Perceived Productivity*. *Lighting Overall* had the weakest correlation for all three variables.

There is a significant difference between the buildings that were classed as 'sustainable' compared to the more conventional buildings. The sustainable buildings scored higher for most variables. This could indicate that for buildings that are considered to be sustainable, designers also aim to make the indoor environment more comfortable, healthy, and productive for the occupants. However, there are a lot of variables associated with the buildings, which could also be having an effect, such as the age of the buildings.

7.1.4 Exploring benchmarks and incorporation methods

The fourth objective was to use the preceding research and the data collected to explore methods for determining benchmarks for NZ commercial buildings in terms of user perceptions.

Several methods of benchmarking were explored over two buildings to determine how they would work with the questionnaire data provided.

These options explored the data by considering a comparison of the data after the results were summarised (absolute benchmarks) and direct comparisons (relative benchmarks). Absolute benchmarks have the advantage of being able to be assessed without the need for a representative sample, while the relative benchmarks gave an assessment of how each variable rated to a larger dataset. Summarising techniques were also explored in terms of each score individually, and the mean score for a question. There were large differences in the results from the way the scores were summarised, and benchmarks produced. This reinforces the importance that standardised methods are used.

It was deemed that it was entirely appropriate and feasible to establish benchmarks from the questionnaire data received, and to incorporate them into building rating tools. There are a variety of forms these benchmarks can take, with positives and negatives for each.

The pilot database is a limited sample of the building stock in New Zealand. Due to this, it is difficult to make conclusive statements about the best method, as the methods have not been tested thoroughly and explored with a representative sample. In terms of which methods are the most promising, they each have their own advantages and disadvantages. The percentage satisfied method, however, which is the method currently incorporated in rating tools, would be a suggested method to explore further for New Zealand as it relates directly to the occupants, has minimal reliance on the influence of the researcher, and has an attainable 100% point.

7.2 Other key factors

The focus of the benchmarks studied in this report was the indoor environmental quality. The indoor environment can have a substantial effect on the occupants that spend a large amount of time in the space, but is not the only factor that can have an effect. The questionnaire also includes how the occupants rate the indoor environmental quality in terms of having effect on their health, and productivity, as well as factors relating to design and needs. Although such matters were not included in the specific building rating tools mentioned earlier, it has been argued (Baird, 2009, p. 1083) that they should be included. The questionnaire based methods studied shows that several of these other variables have been popular with clients.

'Benchmarks our clients find most useful are overall satisfaction, meeting behaviour, quantities of personal storage, and self-assessed productivity' (Alexi Marmot Associates, 2008, p. 1).

The Japanese building rating tool CASBEE has been explored by Takai et al (2008) on its application to existing buildings. They recommend the inclusion of an assessment of productivity.

These previous studies perhaps suggests that a more wide ranging group of variables should be included in any benchmarks that are established in the assessment of users' perceptions of buildings.

7.3 Future work

- As it has been determined that the pilot database is not representative of the building stock, a larger sample needs to be compiled for any such benchmarks. This could be sourced from the BEES data, and would simultaneously provide a profile of the NZ building stock according to users' perceptions.
- There is a significant difference between the buildings classed as 'sustainable' and the more conventional buildings in the pilot database. When a larger database is sourced this would be able to be confirmed with a representative sample, and similar tests will be able to be carried out for the variables classified for the different buildings, such as floor area, ventilation, and location.
- Investigation is needed for creating benchmarks for other sets of variables in the questionnaire, such as satisfaction, operation and control. Separate weightings will have to be determined for these as well as exploring how they could be incorporated into the rating tool. A starting point for the proportions of the weightings would be the correlations with the relevant variable.

WORKS CITED

- ABS Consulting. (2008). *An Introduction to the OLS (Overall Liking Score)*. London: ABS Consulting.
- Action Energy. (2000, December). *Energy Consumption Guide 19: Energy Use in Offices*. Retrieved March 25, 2011, from Chartered Institution of Building Services Engineers:
<http://www.cibse.org/pdfs/ECG019.pdf>
- Aizlewood, C., & Dimitroulopoulou, C. (2006). The HOPE Project: The UK Experience. *Indoor and Built Environment: vol.15*, 393-409.
- Alexi Marmot Associates. (2004, May). *AMA Workware toolkit: case study Department of Health office evaluation*. Retrieved May 11, 2011, from Usable Buildings:
<http://www.usablebuildings.co.uk/rp/OutputFiles/PdfFiles/AMADoH.pdf>
- Alexi Marmot Associates. (2008). *WorkWare Nexus*. Retrieved July 21, 2011, from AMA Alexi Marmot Associates: <http://aleximarmot.com/userfiles/file/AMA%20WorkWare%20Nexus%20Issue%201.pdf>
- American Society of Interior Designers. (2005). *Sound Solutions*. Washington: American Society of Interior Designers.
- Aronoff, S., & Kaplan, A. (1995). *Total Workplace Performance: Rethinking the Office Environment*. Ottawa: WDL Publications.
- ASHRAE. (2004). *ANSI/ASHRAE Standard 55-2004 Thermal Environmental Conditions for Human Occupancy*. Atlanta: ASHRAE.
- ASHRAE; USGBC; CIBSE. (2009). *Performance Measurement Protocols for Commercial Buildings*. (M. S. Owen, Ed.) Atlanta: ASHRAE.
- Australian Department of Environment, Climate Change and Water. (2010). *Indoor Environment: Why rate your building?* Retrieved March 30, 2011, from NABERS:
<http://www.nabers.com.au/page.aspx?cid=616&site=2>
- Babbie, E. (2001). *The Practice of Social Research*. Belmont: Wadsworth Thomson Learning.
- Babbie, E. (2008). *The basics of social research*. Australia: Thomson/Wadsworth.
- Bailey, C. (2010). LEED contender? *CIBSE Journal May*, 30-32.
- Baird, G. (2009). Incorporating User Performance Criteria into Building Sustainability Rating Tools (BRSTs) for Buildings in Operation. *Sustainability, 1*, 1069-1086.
- Baird, G. (2010). *Sustainable Buildings in Practice: What the users think*. Abingdon: Routledge.
- Baird, G., & Oosterhoff, H. (2008). Users' perceptions of health in sustainable buildings - worldwide. *Proceedings of CIB-W70 International Conference in Facilities Management*. Edinburgh.
- Baird, G., Christie, L., Ferris, J., Goguel, C., & Oosterhoff, H. (2008). User perceptions and feedback from the 'best' sustainable buildings in the world. *Proceedings of SB08 - the World Sustainable Buildings Conference*. Melbourne.
- Baird, G., Gray, J., Isaacs, N., Kernohan, D., & McIndoe, G. (1996). *Building Evaluation Techniques*. New York: McGraw-Hill.
- Baird, G., Leaman, A., & Thompson, J. (in press). A Comparison of the Performance of Sustainable Buildings with Conventional Buildings from the Point of View of the Occupants. *Energy & Buildings*.
- Baker, M., Keall, M., Au, E. L., & Howden-Chapman, P. (2007). Home is where the heart is - most of the time. *The New Zealand Medical Journal*, 1-4.

- Bannister, P., Munzinger, M., & Bloomfield, C. (2005). *Water Benchmarks for Offices and Public Benchmarks*. ACT: Exergy Australia Pty Ltd.
- Bell, M. (2010, May 14). *Kurtosis And How It Is Calculated By Statistics Software Packages*. Retrieved May 2, 2011, from <http://www.suite101.com/content/kurtosis-and-how-it-is-calculated-by-statistics-software-packages-a235641>
- Bischof, W., & Bullinger, M. (1998). Indoor Conditions and Well-Being: Interim Results from the ProKlima study. *Indoor and Built Environment*, 7, 232-233.
- Bluyssen, P. (2009). *The Indoor Environment Handbook: How to make buildings healthy and comfortable*. London: Earthscan.
- Bluyssen, P. M., & Cox, C. (2002). Indoor environment quality and upgrading of European office buildings. *Energy and Buildings*, 34, 155-162.
- Bluyssen, P. M., Aries, M., & van Dommelen, P. (2011). Comfort of workers in office buildings: The European HOPE project. *Building and Environment*, 46, 280-288.
- Boonstra, C., & Pettersen, T. D. (2003). Tools for environmental assessment of existing buildings. *UNEP Industry and Environment April - September*, 80-83.
- BRE. (2010). *BREEAM Schemes*. Retrieved August 7, 2011, from BREEAM: The world's leading design and assessment method for sustainable buildings: <http://www.breeam.org/>
- Brill, M. (1985). *Using office design to increase productivity*. Buffalo: Workplace Design and Productivity.
- Brown, J. D. (1997, April). *Skewness and Kurtosis*. Retrieved July 8, 2011, from Statistics Corner: Questions and answers about language testing statistics: http://jalt.org/test/bro_1.htm
- Building Use Studies. (2008). *Environment House*. Building Use Studies.
- Building Use Studies. (2009). *Design*. Retrieved July 7, 2011, from BUS Methodology: <http://homepage.mac.com/aleaman2/10283/index.html>
- Building Use Studies. (2009, August 20). *Summary (Overall variables)*. Retrieved July 22, 2011, from BUS methodology: Occupant Survey Graphics: <http://homepage.mac.com/aleaman2/10282/index.html>
- Building Use Studies. (2009). *The Arup Appraise/BUS occupant survey method: details for licensees*. London: Building Use Studies.
- Building Use Studies. (2009, October 16). *Usable Buildings: Feedback and strategy for better building*. Retrieved May 25, 2011, from Usable Buildings: Feedback and strategy for better building: <http://www.usablebuildings.co.uk/>
- Building Use Studies. (2010, April 12). *Office Productivity Network Survey*. Retrieved July 28, 2010, from Usable Buildings: <http://www.usablebuildings.co.uk/fp/OutputFiles/FR3MainText.html>
- Caccavelli, D., & Gugerli, H. (2002). TOBUS - a European diagnosis and decision-making tool for office building upgrading. *Energy and Buildings* 34, 113-119.
- Carlopio, J. R. (1996). Construct Validity of a Physical Work Environment Satisfaction Questionnaire. *Journal of Occupational Health Psychology*, 1(3), 330-344.
- Center for the Built Environment. (2009, March 20). *Thermal Comfort - CBE Survey*. Retrieved March 25, 2011, from Center for the Built Environment: http://www.cbesurvey.org/CBESurvey/Instrument1040/Thermal.asp?locale=en_US&LID=1&PN=Thermal.asp&SID=2489&IID=1040&PID=9&NP=20&UID=588619&PL=x11110001101010101011&Status=1&pmode=undefined&yScale=undefined

- Center for the Built Environment. (2011, March). *Indoor Environmental Quality Survey Information Packet*. Retrieved May 4, 2011, from Center for the Built Environment: http://www.cbe.berkeley.edu/research/pdf_files/SurveyInfoPacket-2011.pdf
- Chung, W., Hui, Y., & Miu Lam, Y. (2006). Benchmarking the energy efficiency of commercial buildings. *Applied Energy* 83, 1-14.
- CIBSE. (2005). *CIBSE Guide B: Heating, Air Conditioning and Refrigeration*. London: The Chartered Institute of Building Services Engineers .
- Clements-Croome, D., & Baizhan, L. (2000). Productivity and Indoor Environment. *Proceedings of Healthy Buildings 2000*, (pp. 629-634). Espoo.
- Cohen, J. W. (1988). *Statistical power analysis for the behavioural sciences* (2nd Edition ed.). Routledge.
- Cohen, R., Bordass, W., & Field, J. (2006). *Fixed and Customised Benchmarks for Building Energy Performance Certificates based on Operational Ratings*. Retrieved May 14, 2010, from EPLabel: http://www.eplabel.org/links/EPLabel_EPIC_paper_final_03Jul06_corr.pdf
- Cole, R. J. (2003). Green Buildings - Reconciling Technological Change and Occupant Expectations. In R. J. Cole, & R. Lorch, *Buildings, Culture and Environment: Informing local and global practices* (pp. 57-82). Oxford: Blackwell Publishing Ltd.
- Cole, R. J. (2005). Building Environmental Assessment Methods: Redefining Intentions. *Building Research and Information* 33(5), 455-467.
- Construction Industry Council. (2011). *DQI: Design Quality Indicator*. Retrieved August 7, 2011, from DQI: Design Quality Indicator: <http://www.dqi.org.uk/website/default.aspa>
- Creative Research Systems. (2011, February 17). *Sample Size Formulas for our Sample Size Calculator*. Retrieved July 8, 2011, from The Survey System: <http://www.surveysystem.com/sample-size-formula.htm>
- de Dear, R. J. (1998). A global database of thermal comfort field experiments. *ASHRAE Transactions*, 1141-1152.
- de Dear, R., Brager, G. S., & Cooper, D. (1997). *Developing an adaptive model of thermal comfort and preference, Final Report, ASHRAE RP-884*. ASHRAE.
- DeCarlo, L. T. (1997). On the Meaning and Use of Kurtosis. *Psychological Methods* vol. 2 no. 3, 292-307.
- Dorgan, C. E., & Dorgan, C. B. (2006). Assessment of link between productivity and indoor air quality. In D. Clements-Croome, *Creating the productive workplace* (2nd Edition ed., pp. 113-135). Oxon: Taylor & Francis.
- Dwyer, T. (2006). Comfort for productivity in offices. *Building Services Journal*, No. 6, 89-91.
- Dziegielewski, B. (2000). *Commercial and Institutional End Uses of Water*. USA: American Water Works Association.
- Fellows, R., & Liu, A. (2003). *Research Methods for Construction*. Malden: Blackwell Science, Inc.
- Fisk, W. (2002). How IEQ Affects Health, Productivity. *ASHRAE Journal*, 56-60.
- Frontczak, M., & Wargocki, P. (2010). Literature survey on how different factors influence human comfort in indoor environments. *Building and Environment*, 1-16.
- Fullbrook, D., Jackson, Q., & Finlay, G. (2006). *Value Case for Sustainable Building in New Zealand*. Wellington: Ministry for the Environment.

- Gann, D. M., Salter, A. J., & Whyte, J. K. (2003). Design Quality Indicator as a tool for thinking. *Building Research & Information*, 31(5), 318-333.
- Hernandez, P., Burke, K., & Lewis, O. (2008). Development of energy performance benchmarks and building energy ratings for non-domestic buildings: An example for Irish primary schools. *Energy and Buildings* 40, 249-254.
- Higher Education Funding Council. (2006). *Guide to Post Occupancy Evaluation*. Higher Education Funding Council.
- Hinton, P. R., Brownlow, C., McMurray, I., & Cozens, B. (2004). *SPSS Explained*. East Sussex: Routledge.
- Humphreys, M. A. (2005). Quantifying occupant comfort: are combined indices of the indoor environment practicable? *Building Research and Information*, 33(4), 317-325.
- Hyams, D. (2004). The Architect's Perspective. In S. Roaf, *Closing the Loop: Benchmarks for Sustainable Buildings* (pp. 19-25). London: RIBA Enterprises Ltd.
- IBM. (2008, April 10). SPSS 16.0 for Windows. New York, United States.
- Isaacs, N. (2008). Energy Use Research Turns Non-Residential. *BUILD February/March*, 44-45.
- Isaacs, N. (2010, February/March). BEES buzzing. *BUILD* 116, pp. 70-71.
- Isaacs, N., Saville-Smith, K., Bishop, R., Camilleri, M., Jowett, J., Hills, A., et al. (2009). *Building Energy End-Use Study (BEES) Years 1 & 2*. Wellington: BRANZ.
- Julien, A. (2009). Assessing the assessor: BREEAM vs LEED. *Sustain Magazine Vol 9 No 6*, 30-33.
- Juniper, E. (2009). Validated questionnaires should not be modified. *European Respiratory Journal*, 34(5), 1015-1017.
- Leaman, A. (1997). Probe 10: Occupancy Survey Analysis. *Building Services Journal*, 37-41.
- Leaman, A. (2003). User Needs and Expectations. In R. J. Cole, & R. Lorch, *Building, Culture and Environment: Informing local and global practices* (pp. 154-176). Oxford: Blackwell Publishing Ltd.
- Leaman, A. (2004). Post-occupancy Evaluation. In S. Roaf, *Closing the Loop: Benchmarks for sustainable buildings* (pp. 491-518). London: RIBA Enterprises Ltd.
- Leaman, A. (2005). *Research Methods*. Retrieved October 15, 2010, from Usable Buildings: <http://www.usablebuildings.co.uk/WebGuidePDFs/ResearchMethods.pdf>
- Leaman, A. (2010, September 28). BUS Summary Indices. (C. Dykes, Interviewer)
- Leaman, A., & Bordass, B. (2001). Assessing building performance in use 4: the Probe occupant surveys and their implications. *Building Research & Information* 29(2), 129-143.
- Leaman, A., & Bordass, B. (2007, November). Are users more tolerant of 'green' buildings. *Building Research & Information*, pp. 662-673.
- Leaman, A., Stevenson, F., & Bordass, B. (2010, August 24). Building evaluation: practice and principles. *Building Research & Information*, 38(5), 564-577.
- Leaman, A., Thomas, L., & Vandenberg, M. (2007, November). 'Green' Buildings: What Australian Building Users are Saying. *EcoLibrium*, 22-30.
- Levermore, G. J. (1994). Occupants' assessments of indoor environments: Questionnaire and rating scale methods. *Building Services Engineering Research and Technology*, 15(2), 113-118.
- London Hazards Centre. (1995, June). *Air, Light and Temperature*. Retrieved June 29, 2011, from London Hazards Centre Factsheet: <http://www.lhc.org.uk/members/pubs/factsht/47fact.htm>

- Marriage, G. (2010). *Aorangi House - Wellington's Recycled Building*. Wellington: Sustainable Buildings 2010 New Zealand.
- McCartney, K. J., & Humphreys, M. A. (2002). Thermal comfort and productivity. *Indoor Air 2002*, (pp. 822-827). Mexico City.
- Meir, I. M., Garb, Y., Jiao, D., & Cicelsky, A. (2009). Post-Occupancy Evaluation: An Inevitable Step Toward Sustainability. *Advanced in Building Energy Research*, 3, 189-220.
- Mendenhall, W., Beaver, R. J., & Beaver, B. M. (2009). *Introduction to Probability and Statistics*. Belmont: Brooks/Cole, Cengage Learning.
- Miller, G. A. (1956, March). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, 63(2), 81-97.
- Moser, C., & Kalton, G. (1993). *Survey methods in social investigation*. London: Dartmouth.
- National Research Council Canada. (2010, November 26). *Cost-effective Open-Plan Environments (COPE) Project*. Retrieved August 7, 2011, from National Research Council Canada: <http://www.nrc-cnrc.gc.ca/eng/projects/irc/cope.html>
- Nicol, F., & Roaf, S. (2005). Post-occupancy evaluation and field studies of thermal comfort. *Building Research & Information*, 33(4), 338-346.
- NZ Department of Labour. (2009). *Occupational Health Tools*. NZ Department of Labour.
- NZGBC. (2006, June 28). *Ministry for the Environment*. Retrieved April 21, 2011, from Green Building Assessment Tool Research Project: Final Report: <http://www.mfe.govt.nz/publications/sus-dev/green-building-assessment-tool-project-jun06/green-building-assessment-tool-project-jun06.pdf>
- NZGBC. (2008). *Introduction to Rating Tools*. Retrieved July 25, 2011, from Green Star New Zealand: <http://www.nzgbc.org.nz/main/greenstar/elaboration/ratingtools>
- Oesterle, E., Rolf-Pieter, L., Lutz, M., & Heusler, W. (2001). *Double-Skin Facades - Integrated Planning*. Prestel Verlag.
- Oseland, N. (1999). *Environmental factors affecting office worker performance: a review of evidence*. London: Chartered Institute of Building Services Engineers.
- Oseland, N. (2007). *British Council for Offices Guide to Post-Occupancy Evaluation*. London: British Council for Offices.
- Ott, L., & Longnecker, M. (2010). *An Introduction to Statistical Methods and Data Analysis, Sixth Edition*. Belmont: Brooks/Cole, Cengage Learning.
- Oxford University Press. (2010, April). *Benchmark: Oxford Dictionaries Online*. Retrieved November 26, 2010, from Oxford Dictionaries Online: http://english.oxforddictionaries.com/view/entry/m_en_gb0071800#m_en_gb0071800
- Oxford University Press. (2010, April). *Perception*. Retrieved October 18, 2010, from Oxford Dictionaries: http://english.oxforddictionaries.com/view/entry/m_en_gb0618760#m_en_gb0618760
- Pallant, J. (2005). *SPSS survival manual: a step by step guide to data analysis using SPSS*. Crows Nest, NSW: Allen & Unwin.
- Peck, R., Olsen, C., & Devore, J. (2009). *Introduction to Statistics and Data Analysis*. Belmont: Brooks/Cole, Cengage Learning.

- Peretti, C., & Schiavon, S. (2011, January 6). *Indoor environmental quality surveys. A brief literature review*. Retrieved June 20, 2011, from eScholarship University of California: <http://escholarship.org/uc/item/2p08233r>
- Perez-Lombard, L., Ortiz, J., Gonzalez, R., & Maestre, I. (2009). A review of benchmarking, rating and labelling concepts within the framework of building energy certification schemes. *Energy and Buildings* 41, 272-278.
- Prasad, S. (2004). Clarifying intentions: the design quality indicator. *Building Research & Information*, 32(6), 548-551.
- Preiser, W. (2010, October 18). Feedback, feedforward and control: post-occupancy evaluation to the rescue. *Building Research & Information*, 29(6), 456-459.
- Roaf, S. (2004). *Closing the Loop: Benchmarks for sustainable buildings*. London: RIBA Enterprises Ltd.
- Roelofsen, P. (2002). The impact of office environments on employee performance: The design of the workplace as a strategy for productivity enhancement. *Journal of Facilities Management*, Vol.1 No. 3, 247-264.
- Roulet, C.-A., Flourentzou, F., Foradini, F., Bluysen, P., Cox, C., & Aizlewood, C. (2006). Multicriteria analysis of health, comfort and energy efficiency in buildings. *Building Research & Information*, 34(5), 475-482.
- Salkind, N. (2005). *Statistics for People Who (Think They) Hate Statistics: 2nd Edition*. USA: Sage Publications.
- Santamouris, M., Mihalakakou, G., Patargias, P., Gaitani, N., Sfakianaki, K., Papaglastra, M., et al. (2007). Using intelligent clustering techniques to classify the energy performance of school buildings. *Energy and Buildings* 39, 45-51.
- Schellen, L., van Marken Lichtenbelt, W., de Wit, M., Loomans, M., Frijns, A., & Toftum, J. (2008). Thermal comfort, physiological responses and performance during exposure to a moderate temperature drift. *Indoor Air 2008*. Copenhagen.
- Smith, J. (2008). *[Masters Thesis] Implementation of a Building Sustainability Rating Tool: A Survey of the New Zealand Building Industry*. Wellington: Victoria University of Wellington.
- Stokols, D., & Scharf, T. (1990). Developing standardised tools for assessing employees' ratings of facility performance. In G. Davis, & F. Ventre, *Performance of buildings and serviceability of facilities* (pp. 55-79). Philadelphia: American Society for Testing and Materials.
- Sustainable Built Environments, Centre for Design at RMIT University. (2007). *ESD Design Guide: Office and Public Buildings* (Vol. Edition 3). Melbourne: Australian Department of the Environment and Water Resources.
- Swanke Hayden Connell Architects. (2005). *OPN Workplace Evaluation Survey*. Retrieved March 23, 2011, from Office Productivity Network: <http://www.officeproductivity.co.uk/files/OPN%20Survey.pdf>
- Swanke Hayden Connell Architects. (2005). *POE Methodology*. Retrieved March 23, 2011, from Office Productivity Network: <http://www.officeproductivity.co.uk/documents/POEmethodology201205.doc>
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4, illustrated ed.). Michigan: Allyn and Bacon.
- Takai, H., Murakami, S., Ikaga, T., Ito, M., & Sakai, T. (2008). Three Studies on the Promotion of Assessment Tools and Market Transformation: The Case of CASBEE. *Proceedings of the World Sustainable Building Conference SB08*, (pp. 1500-1507). Melbourne.
- U.S. Environmental Protection Agency. (2003). *A Standardised EPA Protocol for Characterising Indoor Air Quality in Large Office Buildings*. Washnigton, D.C.: U.S. Environmental Protection Agency.

- Usable Buildings Trust. (2011, 4 February). *Usable Buildings*. Retrieved May 25, 2011, from Usable Buildings: <http://www.usablebuildings.co.uk/>
- USGBC. (2009). *LEED 2009 For New Construction and Major Renovations*. Retrieved May 2, 2011, from U.S. Green Building Council: <http://www.usgbc.org/ShowFile.aspx?DocumentID=5546>
- USGBC. (2011, February). *LEED 2009 For Existing Buildings Operations and Maintenance*. Retrieved May 2, 2011, from U.S. Green Building Council: <http://www.usgbc.org/ShowFile.aspx?DocumentID=8876>
- USGBC. (2011). *LEED: How to achieve certification*. Retrieved April 21, 2011, from U.S. Green Building Council: <http://www.usgbc.org/DisplayPage.aspx?CMSPageID=1991>
- Volker, L., & van der Voordt, D. J. (2005). An integral tool for the diagnostic evaluation of non-territorial offices. In B. Martens, & A. G. Keul, *Designing Social Innovation. Planning, Building, Evaluating* (pp. 241-250). Gottingen: Hogrefe & Huber Publishers.
- Volker, L., & van der Voort, T. J. (2009). Measuring employee satisfaction in new offices - the WODI toolkit. *Journal of Facilities Management*, 7(3), 181-197.
- Waitakere Properties Limited. (2007). *Waitakere Central*. Auckland: Trends.
- World Health Organisation. (2011). *World Health Organisation*. Retrieved July 26, 2011, from World Health Organisation: <http://www.who.int/en/>

APPENDICES

Appendix A: Questionnaire based methods

Section 2.3: Examples of questionnaires used, and Section 3.3: Benchmarks of people's perceptions, assess a sample of 20 questionnaire based methodologies. The main factors have been summarised in two tables, in these sections. There are multiple sources for the assessment but the tables are largely based on Peretti and Schiavon (2011), and Usable Buildings Trust (2011). Table 50 provides additional information on the questionnaire based methods.

Table 50: Examples of questionnaire based methods

	Survey Name	Objectives	Questionnaire Structure	Physical Measurements
1	BOSTI (Buffalo Organisation for Social and Technological Innovation) Survey	Using office design to increase productivity	Workspace & layout, Ambient conditions, workspace design, privacy, interaction & communication patterns, Workers & their jobs, Organisations.	Not Performed
2	BUS Occupant Survey	Assess how well buildings work, get feedback on occupant needs and perceptions, improve services to occupants.	background, overall building, internal environment, travel to work	Not Performed
3	Physical Work Environment Satisfaction Questionnaire (PWESQ)	Effects of the physical environment on employee behaviour and attitudes. Survey on satisfaction with the physical environment and job satisfaction.	Questionnaire is composed of 42 items.	Not Performed
4	REF (Ratings of Environmental Features)	Research strategies for evaluating facility design, occupant productivity and organizational effectiveness. Suggest specific organizational and environmental design strategies.	Basic survey:24 items, Complete survey:48 items	Not Performed
5	CWRE (Checklist of Work Related Experiences)	Developing standardised tools for assessing employees' rating of facility performance	Lists 30-35 different work related experiences and respondents answer in terms of how often they experience each encounter.	Not Performed
6	AMA WorkWare	To support client and users needing to improve the effectiveness and efficiency of existing space or achieve a better future project.	Lists different factors and gets the occupants to rate them.	Not Performed
7	OLS (Overall Liking Score)	Obtain information on occupations likes and concerns, to identify successful features of a building, or as a KPI for maintenance and other facilities management services.	5 Sections: Background, Summer and winter comfort, factors relation to interior environment, sickness symptoms, stress.	Not Performed
8	Building Assessment Survey and Evaluation (BASE)	Occupant perceptions of IAQ and health symptoms	33 questions and additional space for comments	Mobile cart: CO ₂ , temperature, RH and supply air delivery. Real time monitors: CO, CO ₂ , temperature, RH, VOCs, PM _{2.5} , PM ₁₀
9	ProKlima	Contribution of the indoor climate, energy concept and psychological factors to the illness symptoms and thermal comfort.	Physical perceptions, health, well-being, indoor climate (including odours, light, noise, temperature), satisfaction with work and personal statistics.	Indoor air quality, noises, thermal comfort parameters and light were measured. Temperatures and humidity at one height over a period of 15 mins in each building.
10	ASHRAE RP-884	Develop an adaptive thermal comfort standard for ASHRAE.	Background questionnaire and thermal comfort questionnaire	Clothing insulation, metabolic rate, outdoor meteorological conditions, indoor air and mean radiant temperature, air speed, indoor humidity
11	CBE Survey (Centre for the Built Environment)	Evaluation of building technologies and performance, quality benchmarking, diagnosis of problems	Core Survey (about 60 questions). Custom modules can be added to address issues not covered in the core questions.	Dependent on which project the measurements were associated
12	SCATS (Smarts Control and Thermal Comfort)	Correlation between comfort temperatures and indoor/outdoor temperatures, behavioural analyses. Developing an adaptive control algorithm for Europe.	Transverse questionnaire: 16 questions. Longitudinal questionnaire: 5 questions	CO ₂ concentration, globe temperature, air temperature, relative humidity, illuminance, air velocity, noise level, meteorological situations for outdoor parameters (Simultaneous with Transverse Questionnaire)
13	HEFC(Higher Education Funding Council) methodology	Develop a toolkit on good practice guidance. To encourage good building design by allowing others to learn from the experience of others.	General background including time and location in the buildings, rating of the building generally, and then location specific including IEQ	Dependent on which project the measurements were associated

14	DQI (Design Quality Indicator)	Process for evaluating functionality, build quality and impact of buildings.	100 statements that measure success factors. Three sections: Functionality; Building Quality; Impact, Ten Sub-sections	Not Performed
15	OPN (Office Productivity Network)	To determine how well the office environment supports individual work activities and productivity	Six sections: Office facilities; Environmental Conditions; Work duties; Downtime factors; Workplace activities; Personal details.	Not Performed
16	TOBUS (Tool for Selecting Office Building Upgrading Solutions)	Computer software designed to present how a building is performing and how best to improve it.	IEQ	Not Performed (spot measurements)
17	COPE (Cost-effective Open Plan Environments)	Evaluation of indoor environment satisfaction of occupants. How the physical environment influences organizational outcomes (job satisfaction, absenteeism, turnover, productivity)	5 demographic questions, 18 individual Environmental Features Ratings, 2 Overall Environment Satisfaction questions, 2 questions on Job Satisfaction.	Physical measurements of each participant's workstation. Cart+chair system and acoustics measurements at night (illumination, air velocity, CO, CO ₂ , THC, CH ₄ , TVOC, temperature, relative humidity)
18	HOPE (Health Optimisation Protocol for Energy-efficient Buildings)	SBS research, benchmarking of healthy and energy efficient buildings.	5 comfort items, 7 SBS items and 12 illness indicators	yes for 32 buildings (and energy use data)
19	WEDI (Work Environment Diagnosis Instrument)	Toolkit to measure employee satisfaction and perceived labour productivity as affected by different workplace strategies	39 questions. Organisation and work; building; direct work environment; privacy; workplace; concentration; communication; archive; IT; indoor climate; external services; and perceived work productivity	Not Performed
20	RPM (Remote Performance Measurement)	Evaluation of IEQ satisfaction, health conditions and personal control by occupants. Characterisation of occupant perceptions and symptoms.	Background questionnaire: occupant general perception of the indoor environment. Instant questionnaire: effects on occupants of any intervention performed	Dependent on which project the measurements were associated

Table 51: References for questionnaire examples

1	BOSTI	(Brill, 1985)
2	BUS	(Usable Buildings Trust, 2011)
3	PWESQ	(Carlopio, 1996)
4	REF	(Stokols & Scharf, 1990)
5	CWRE	(Stokols & Scharf, 1990)
6	AMA	(Alexi Marmot Associates, 2004; Alexi Marmot Associates, 2008)
7	OLS	(Levermore, 1994; ABS Consulting, 2008)
8	BASE	(U.S. Environmental Protection Agency, 2003)
9	ProKlima	(Bischof & Bullinger, 1998)
10	ASHRAE	(de Dear, Brager, & Cooper, 1997; de Dear R. J., 1998)
11	CBE	(Center for the Built Environment, 2011)
12	SCATS	(McCartney & Humphreys, 2002; Humphreys, 2005)
13	HEFC	(Higher Education Funding Council, 2006)
14	DQI	(Construction Industry Council, 2011; Prasad, 2004; Gann, Salter, & Whyte, 2003)
15	OPN	(Swanke Hayden Connell Architects, 2005; Swanke Hayden Connell Architects, 2005)
16	TOBUS	(Bluyssen & Cox, 2002; Caccavelli & Gugerli, 2002)
17	COPE	(National Research Council Canada, 2010)
18	HOPE	(Bluyssen, Aries, & van Dommelen, 2011; Aizlewood & Dimitroulopoulou, 2006)
19	WEDI	(Volker & van der Voordt, 2005; Volker & van der Voort, 2009)
20	RPM	(Schellen, van Marken Lichtenbelt, de Wit, Loomans, Frijns, & Toftum, 2008)

Appendix B: BUS rating score method

For the scales with an ideal score of 7:

The ratings of 5 and 1 are straightforward

- 5 corresponds to a 'Better' rating where the building score is above both the confidence limits of the benchmark and the scale midpoint.

- *Example: Design: Rating = 5*

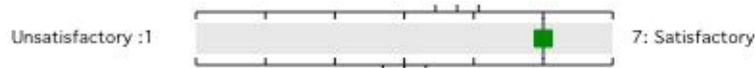


Figure 82: Example of a 5 rating

- 1 corresponds to a 'Worse' rating where the building score is below both the confidence limits of the benchmark and the scale midpoint.

- *Example : Air Overall in Summer: Rating = 1*



Figure 83: Example of a 1 rating

2, 3 and 4 correspond to a rating of 'Similar'

- For a building score that is within both the scale midpoint and the benchmark the variable will get a corresponding score of 3.

- *Example: Air Overall in Summer: Rating = 3*



Figure 84: Example of a 3 rating

- For a building score that is higher than either the scale midpoint or the benchmark but still within the confidence limits of one of them will get a corresponding score of 4.

- *Example: Lighting Overall: Rating = 4*

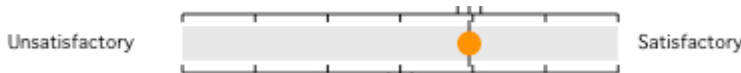


Figure 85: Example of a 4 rating

- For a building score that is lower than either the scale midpoint or the benchmark but still within the confidence limits of one of them will get a corresponding score of 2.

- *Example: Design: Rating = 2*



Figure 86: Example of a 2 rating

Examples are from the NZ buildings in the pilot database (Building Use Studies, 2009).

Appendix C: Adequate sample size calculation

The sample size calculation can be used to determine how many people you need to survey in order to get results that reflect the target population as precisely as needed. The formula was sourced from Creative Research Systems (2011).

Desired Sample Size

Equation 11: Sample size calculation

$$ss = \frac{Z^2 * (p) * (1 - p)}{c^2}$$

Where:

ss = sample size

Z = z value (e.g. 1.96 for 95% confidence level)

p = percentage picking a choice, expressed as a decimal (e.g. 0.5 is worst case scenario and used for sample size needed)

c = margin of error (confidence interval), expressed as a decimal (e.g. 0.05 = ± 5%)

$$ss = \frac{Z^2 * (p) * (1 - p)}{c^2}$$

$$ss = \frac{1.96^2 * (0.5) * (1 - 0.5)}{0.05^2}$$

$$ss = 384$$

As the building stock has a finite population there is an additional correction that is added to the equation. According to BEES, the number of office buildings in NZ is 6,536 so this will be used as the *pop* value.

$$new\ ss = \frac{ss}{1 + \frac{ss - 1}{pop}}$$

$$new\ ss = \frac{384}{1 + \frac{384 - 1}{6536}}$$

$$new\ ss = 363$$

The calculation can also be rearranged to determine the level of precision you have in an existing sample (Creative Research Systems, 2011). Therefore using the pilot database sample size of 30 buildings and using a confidence level of 95% the margin of error (confidence interval) can be determined.

c = margin of error

$$c = \sqrt{\frac{Z^2 * p * (1 - p)}{30}}$$

$$c = \sqrt{\frac{0.9604}{30}}$$

$$c = 0.1785$$

Confidence level = 95%

Margin of error = $\pm 18\%$

Appendix D: Normality tests

A Kolmogorov-Smirnov test was carried out on the 16 summary variables through SPSS. The resultant table is shown below in Table 52.

Table 52: Kolmogorov-Smirnov test results

	N	Normal Parameters ^a		Most Extreme Differences			Kolmogorov-Smirnov Z	Asymp. Sig. (2-tailed)
		Mean	Std. Deviation	Absolute	Positive	Negative		
Design	30	4.8557	.74692	.111	.106	-.111	.608	.854
Needs	30	4.9190	.66252	.097	.069	-.097	.529	.942
Space	21	4.7824	.60375	.143	.143	-.130	.656	.782
Image	30	4.9140	1.13182	.124	.088	-.124	.680	.744
Meeting Rooms	30	4.5227	.80670	.122	.113	-.122	.667	.765
Storage	30	4.0900	0.62517	.098	.098	-.088	.538	.934
Furniture	30	5.1847	.43278	.125	.125	-.087	.684	.737
Temperature in Winter: Overall	30	4.4973	.45083	.066	.066	-.056	.363	.999
Air in Winter- Overall	23	4.3913	.53908	.132	.132	-.070	.634	.816
Temperature in Summer: Overall	30	4.2923	.49155	.101	.101	-.072	.553	.920
Air in Summer- Overall	23	4.2465	.52100	.167	.167	-.088	.801	.542
Noise overall	30	4.3547	.49000	.102	.102	-.089	.561	.912
Lighting overall	30	5.0090	.33561	.143	.118	-.143	.785	.569
Overall comfort	30	4.6993	.58959	.107	.107	-.060	.585	.883
Perceived Productivity	30	-.9433	5.53057	.104	.104	-.082	.570	.901
Health	30	3.7250	.53146	.167	.167	-.099	.915	.372

a. Test distribution is Normal.

Appendix E: Correlations

Appendix E.1: Individual variables correlations

As mentioned in **Section 5.4: Questionnaire variables**, the 16 summary variables were correlated with each other to determine any relationships between them. Those for *Overall Comfort*, *Health* and *Perceived Productivity* were focussed on in the main body of the thesis. Table 53 displays the corresponding Pearson Correlation Coefficients for all 16 summary variables.

Table 53: Correlation of summary variables

Correlations																
	D	N	S	I	MR	St	F	TW	AW	TS	AS	N	L	C	P	H
Design	-	.90	.87	.93	.70	.74	.81	.82	.71	.62	.52	.56	.49	.92	.81	.68
Needs	.90	-	.88	.75	.68	.75	.76	.75	.72	.67	.68	.63	.64	.89	.84	.65
Space	.87	.88	-	.84	.86	.64	.73	.76	.66	.57	.59	.61	.49	.82	.78	.64
Image	.93	.75	.84	-	.64	.54	.73	.74	.56	.51	.36	.47	.35	.81	.71	.60
Meeting Rooms	.70	.68	.86	.64	-	.69	.64	.65	.67	.49	.52	.47	.57	.73	.70	.44
Storage	.74	.75	.64	.54	.69	-	.77	.69	.72	.64	.64	.69	.64	.85	.82	.61
Furniture	.81	.76	.73	.73	.64	.77	-	.60	.51	.55	.37	.59	.44	.78	.81	.56
Temp Winter	.82	.75	.76	.74	.65	.69	.60	-	.94	.72	.71	.43	.48	.87	.76	.64
Air in Winter	.71	.72	.66	.56	.67	.72	.51	.94	-	.65	.65	.40	.56	.80	.78	.68
Temp Summer	.62	.67	.57	.51	.49	.64	.55	.72	.65	-	.92	.46	.59	.82	.76	.66
Air in Summer	.52	.68	.59	.36	.52	.64	.37	.71	.66	.92	-	.53	.62	.75	.74	.75
Noise	.58	.63	.61	.47	.47	.69	.59	.43	.40	.46	.53	-	.52	.69	.71	.43
Lighting	.49	.64	.49	.35	.57	.64	.44	.48	.56	.59	.62	.52	-	.63	.71	.44
Overall comfort	.92	.89	.82	.81	.73	.85	.78	.87	.80	.82	.75	.69	.63	-	.90	.74
Productivity	.81	.84	.78	.71	.70	.82	.81	.76	.78	.76	.74	.71	.71	.90	-	.71
Health	.68	.65	.64	.60	.44	.61	.56	.64	.68	.66	.75	.43	.44	.74	.71	-

Appendix E.2: Correlations of variables to indices

Table 54: Correlations of indices and summary variables

Correlations						
	UK Comfort Index	NZ Comfort Index	UK Satisfaction Index	NZ Satisfaction Index	UK Summary Index	NZ Summary Index
Overall comfort	.952	.948	-	-	.970	.968
Lighting	.741	.744	-	-	.717	.717
Noise	.734	.709	-	-	.741	.730
Temp in Winter	.898	.906	-	-	.872	.875
Temp in Summer	.864	.867	-	-	.816	.814
Air in Winter	.846	.862	-	-	.827	.834
Air in Summer	.878	.876	-	-	.811	.809
Design	-	-	.910	.912	.880	.881
Needs	-	-	.898	.901	.927	.928
Health	-	-	.890	.887	.877	.877
Productivity	-	-	.899	.900	.954	.955

Appendix E.3: Correlations to BUS rating score: selected variable method

Table 55: Correlations of BUS rating score: selected variable method and summary variables

Correlations				
	UK rating score	UK Percentage	NZ rating score	NZ Percentage
Design	.864	.909	.883	.915
Needs	.848	.890	.869	.893
Image	.756	.803	.770	.808
Temp in Winter	.789	.808	.769	.813
Temp in Summer	.683	.746	.753	.784
Noise	.610	.616	.639	.642
Lighting	.487	.555	.600	.629
Overall comfort	.883	.928	.922	.947
Productivity	.768	.838	.878	.887
Health	.598	.683	.649	.713
UK Percentage	.977	-	.981	.976
NZ Percentage	.949	.981	.976	-

Appendix F: Building variables

Appendix F.1: Independent t-tests

The buildings were classified into groups of buildings with sustainable intentions, and building without. The statistics for the two separate groups are shown below in Table 56.

Table 56: Independent samples t-test group statistics

Group Statistics					
	Sustainable or Conventional	N	Mean	Std. Deviation	Std. Error Mean
Design	Sust	13	5.3200	.72320	.20058
	Conv	17	4.5006	.55657	.13499
Needs	Sust	13	5.2446	.63660	.17656
	Conv	17	4.6700	.58274	.14134
Space	Sust	11	5.0791	.52634	.15870
	Conv	10	4.4560	.52559	.16621
Image	Sust	13	5.7754	.89496	.24822
	Conv	17	4.2553	.81072	.19663
Meeting Rooms	Sust	13	4.9969	.59728	.16565
	Conv	17	4.1600	.76777	.18621
Storage	Sust	13	4.3023	.73535	.20395
	Conv	17	3.9276	.48808	.11838
Furniture	Sust	13	5.4215	.46957	.13024
	Conv	17	5.0035	.30602	.07422
Temperature in Winter: Overall	Sust	13	4.7031	.46568	.12916
	Conv	17	4.3400	.38085	.09237
Air in Winter- Overall	Sust	11	4.5709	.61020	.18398
	Conv	12	4.2267	.42532	.12278
Temperature in Summer: Overall	Sust	13	4.5100	.49578	.13751
	Conv	17	4.1259	.43089	.10451
Air in Summer- Overall	Sust	11	4.3773	.59888	.18057
	Conv	12	4.1267	.42902	.12385
Noise overall	Sust	13	4.5223	.51136	.14182
	Conv	17	4.2265	.44584	.10813
Lighting overall	Sust	13	5.1723	.30822	.08549
	Conv	17	4.8841	.30767	.07462
Overall comfort	Sust	13	5.0323	.64357	.17849
	Conv	17	4.4447	.40055	.09715
Perceived Productivity	Sust	13	2.5877	5.33630	1.48002
	Conv	17	-3.6435	4.02557	.97634
Health	Sust	13	3.9992	.59049	.16377
	Conv	17	3.5153	.37764	.09159

The full results from the independent samples t-test are shown below.

Table 57: Independent samples t-test results

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Design	Equal variances assumed	.339	.565	3.511	28	.002	.81941	.23336	.3414	1.2974
	Equal variances not assumed			3.389	21.954	.003	.81941	.24177	.3179	1.3209
Needs	Equal variances assumed	.115	.737	2.572	28	.016	.57462	.22342	.1169	1.0323
	Equal variances not assumed			2.541	24.700	.018	.57462	.22616	.1085	1.0407
Space	Equal variances assumed	.129	.723	2.711	19	.014	.62309	.22982	.1421	1.1041
	Equal variances not assumed			2.711	18.816	.014	.62309	.22980	.1418	1.1044
Image	Equal variances assumed	.049	.826	4.866	28	.000	1.52009	.31238	.8802	2.1599
	Equal variances not assumed			4.800	24.539	.000	1.52009	.31666	.8673	2.1729
Meeting Rooms	Equal variances assumed	.466	.501	3.246	28	.003	.83692	.25784	.3088	1.3651
	Equal variances not assumed			3.358	27.980	.002	.83692	.24923	.3264	1.3475
Storage	Equal variances assumed	2.303	.140	1.677	28	.105	.37466	.22347	-.0831	.8324
	Equal variances not assumed			1.589	19.765	.128	.37466	.23581	-.1176	.8669
Furniture	Equal variances assumed	1.795	.191	2.949	28	.006	.41801	.14175	.1277	.7084
	Equal variances not assumed			2.789	19.517	.012	.41801	.14990	.1048	.7312
Temperature in Winter: Overall	Equal variances assumed	.888	.354	2.350	28	.026	.36308	.15449	.0466	.6795
	Equal variances not assumed			2.287	22.918	.032	.36308	.15879	.0345	.6916
Air in Winter- Overall	Equal variances assumed	2.810	.109	1.581	21	.129	.34424	.21773	-.1086	.7970
	Equal variances not assumed			1.556	17.699	.137	.34424	.22119	-.1210	.8095
Temperature in Summer: Overall	Equal variances assumed	.778	.385	2.267	28	.031	.38412	.16942	.0371	.7312
	Equal variances not assumed			2.224	23.889	.036	.38412	.17271	.0276	.7407
Air in Summer- Overall	Equal variances assumed	1.765	.198	1.161	21	.258	.25061	.21577	-.1981	.6993
	Equal variances not assumed			1.145	18.000	.267	.25061	.21896	-.2094	.7106
Noise overall	Equal variances assumed	.061	.807	1.690	28	.102	.29584	.17502	-.0627	.6544
	Equal variances not assumed			1.659	23.939	.110	.29584	.17834	-.0723	.6640
Lighting overall	Equal variances assumed	.038	.847	2.540	28	.017	.28819	.11345	.0558	.5206
	Equal variances not assumed			2.540	25.953	.017	.28819	.11347	.0549	.5215
Overall comfort	Equal variances assumed	2.590	.119	3.074	28	.005	.58760	.19116	.1960	.9792

	Equal variances not assumed			2.891	18.917	.009	.58760	.20322	.1621	1.0131
Perceived Productivity	Equal variances assumed	2.057	.163	3.650	28	.001	6.23122	1.70695	2.7347	9.7276
	Equal variances not assumed			3.514	21.643	.002	6.23122	1.77305	2.5506	9.9118
Health	Equal variances assumed	3.668	.066	2.733	28	.011	.48394	.17705	.1213	.8466
	Equal variances not assumed			2.579	19.267	.018	.48394	.18764	.0916	.8763

The values for the effect size calculation can be sourced from Table 57 above. The effect size calculation is:

$$Eta\ squared = \frac{t^2}{t^2 + (N1 + N2 - 2)}$$

The t value can be sourced from the column labelled 't'. The values for N1 and N2 are the different sample sizes which can be sourced from Table 56, which in this case is 13 and 17.

Appendix G: Individual building distributions

Appendix G.1: Building 21 – low scoring building

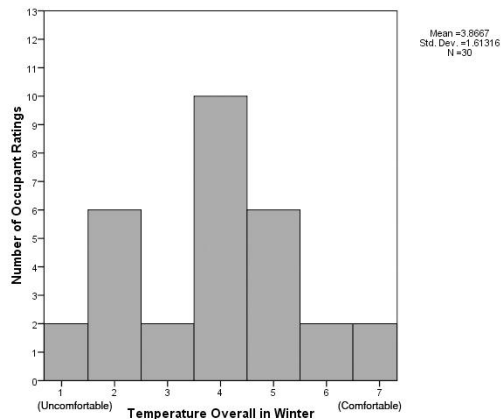


Figure 87: Building 21 - Temperature Overall in Winter

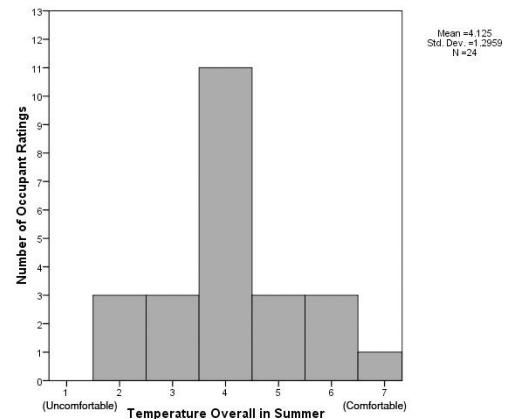


Figure 88: Building 21 - Temperature Overall in Summer

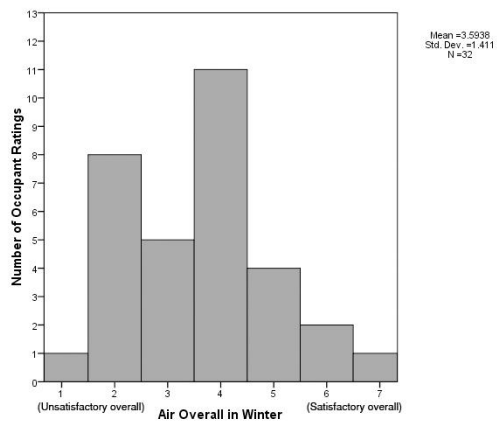


Figure 89: Building 21 - Air Overall in Winter

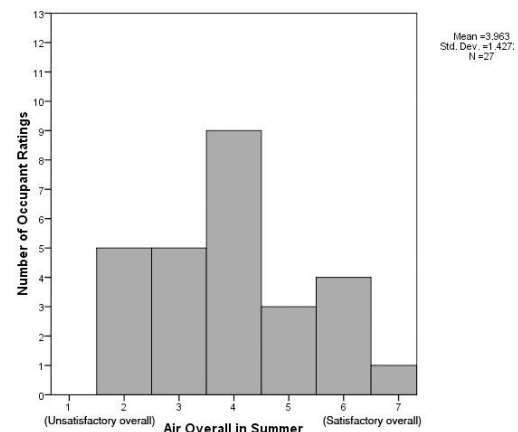


Figure 90: Building 21 - Air Overall in Summer

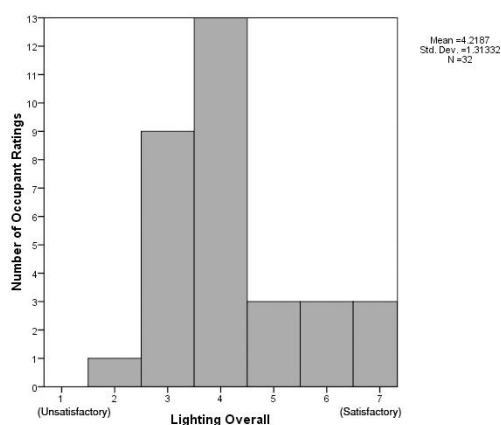


Figure 91: Building 21 - Lighting Overall

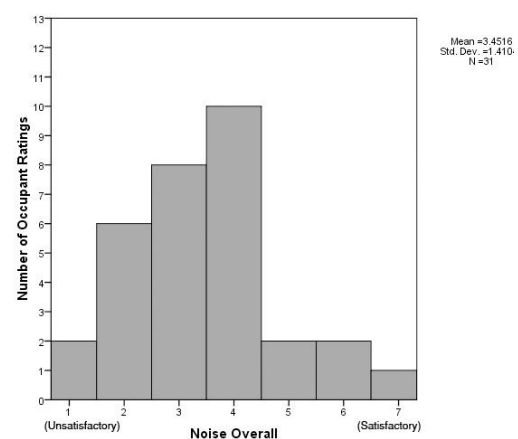


Figure 92: Building 21 - Noise Overall

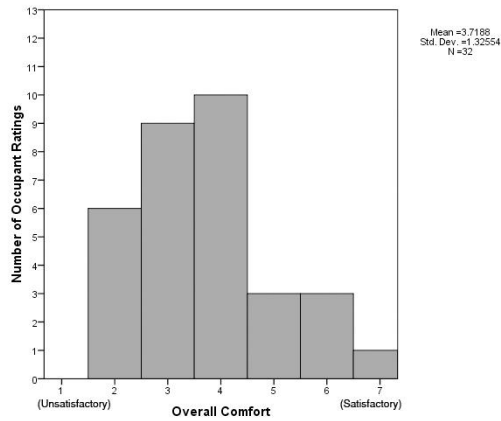


Figure 93: Building 21 - Overall Comfort

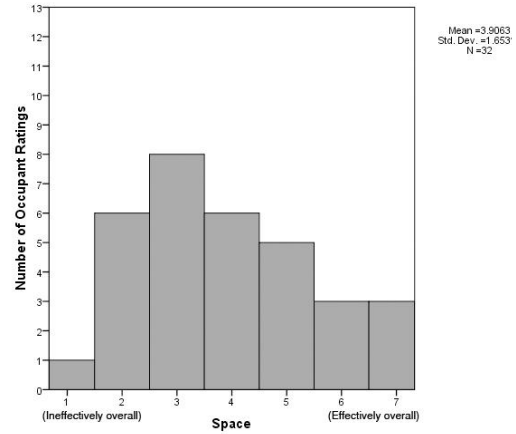


Figure 94: Building 21 - Space

Appendix G.2: Building 29 – high scoring building

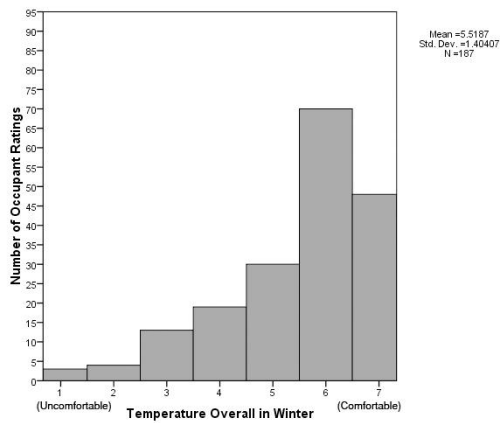


Figure 95: Building 29 - Temperature Overall in Winter

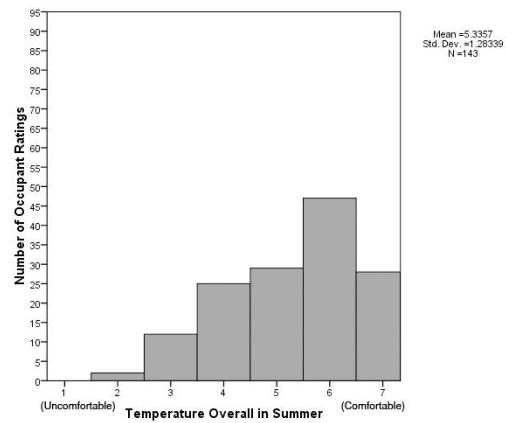


Figure 96: Building 29 - Temperature Overall in Summer

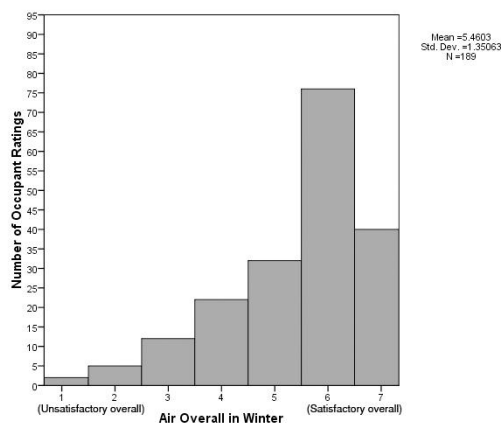


Figure 97: Building 29 - Air Overall in Winter

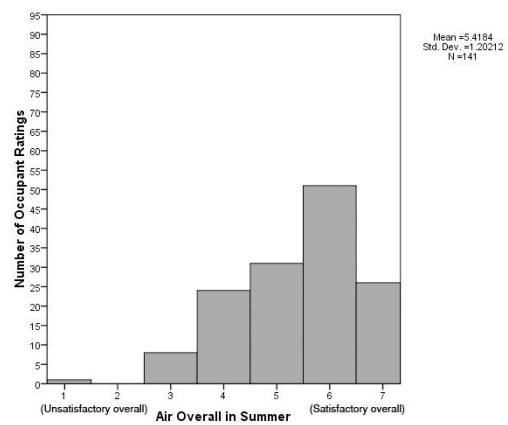


Figure 98: Building 29 - Air Overall in Summer

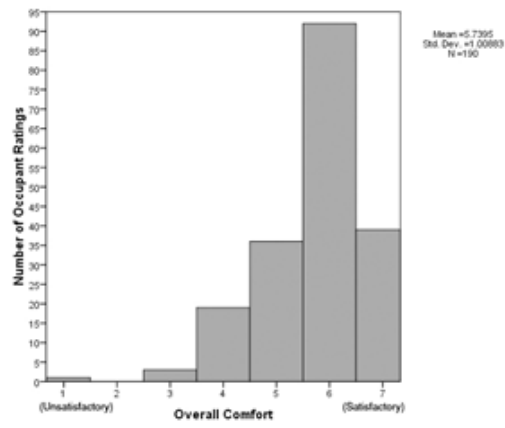


Figure 99: Building 29 - Overall Comfort

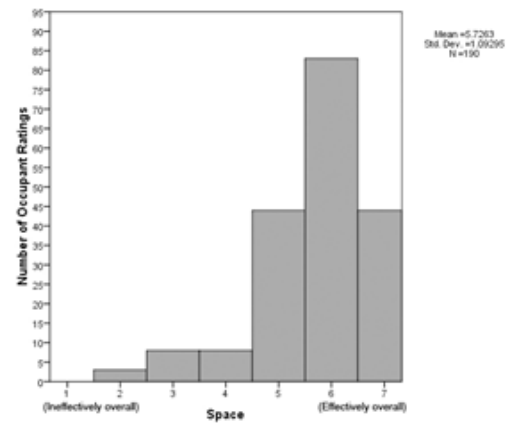


Figure 100: Building 29 - Space

Appendix H: Internal environment variables – distributions

Appendix H.1: Conditions in winter – Temperature and Air Quality

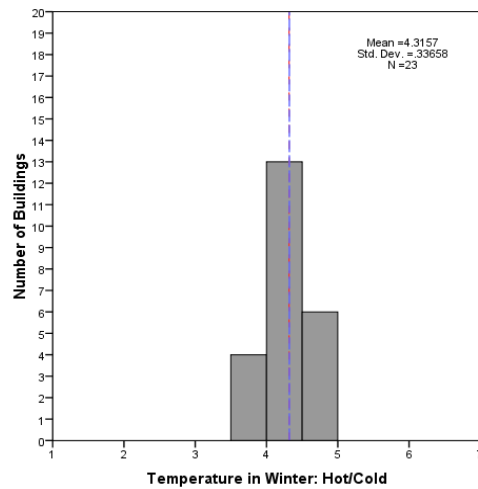


Figure 101: Temperature in Winter: Hot/Cold

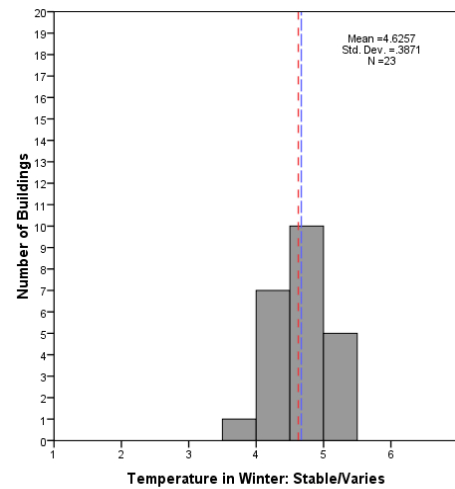


Figure 102: Temperature in Winter: Stable/Varies

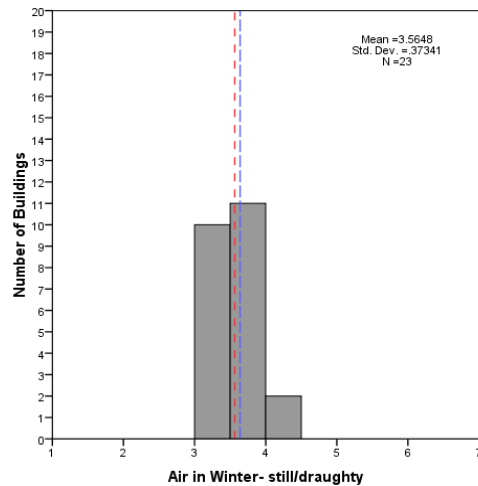


Figure 103: Air in Winter: Still/Draughty

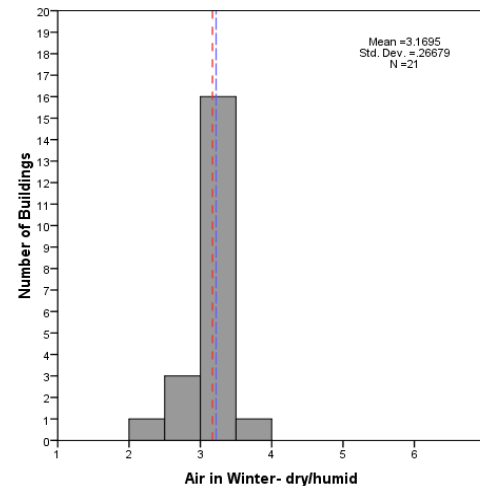


Figure 104: Air in Winter: Dry/Humid

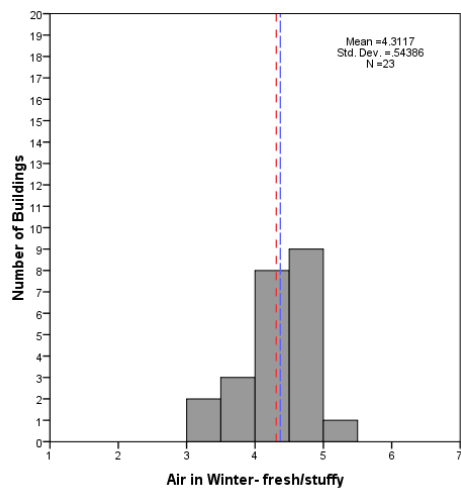


Figure 105: Air in Winter: Fresh/Stuffy

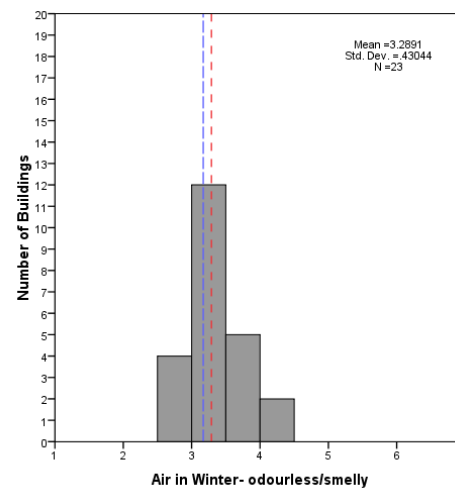


Figure 106: Air in Winter: Odourless/Smelly

Appendix H.2: Conditions in summer – *Temperature and Air Quality*

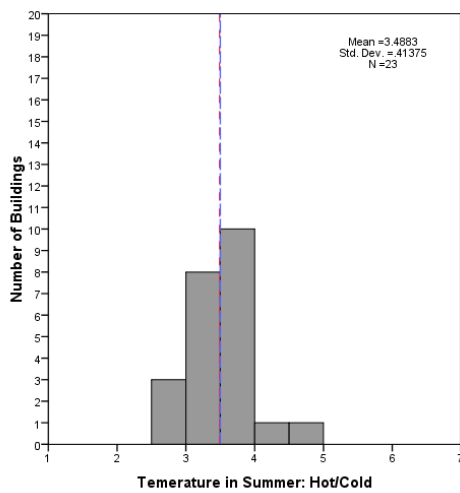


Figure 107: Temperature in Summer: Hot/Cold

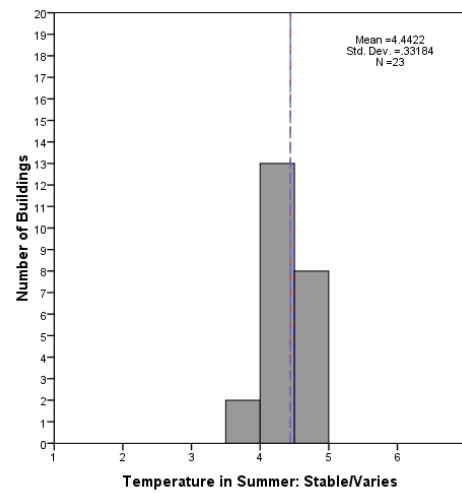


Figure 108: Temperature in Summer: Stable/Varies

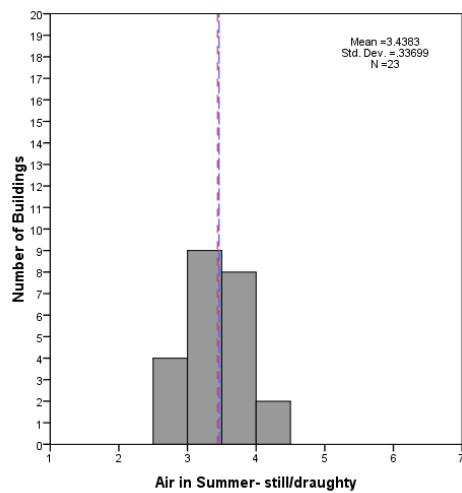


Figure 109: Air in Summer: Still/Draughty

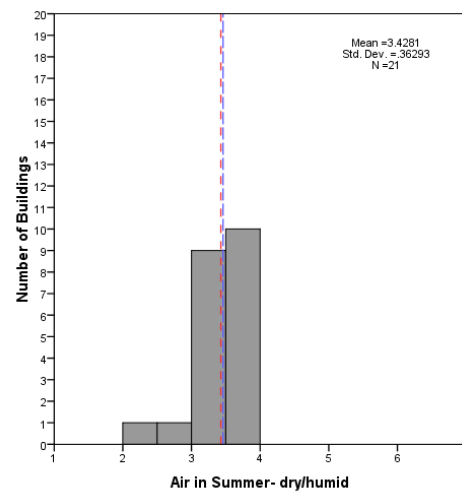


Figure 110: Air in Summer: Dry/Humid

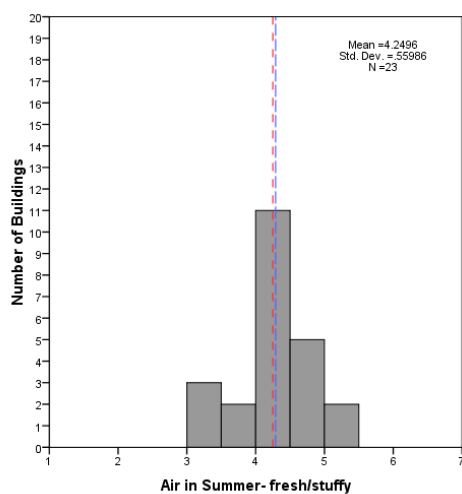


Figure 111: Air in Summer: Fresh/Stuffy

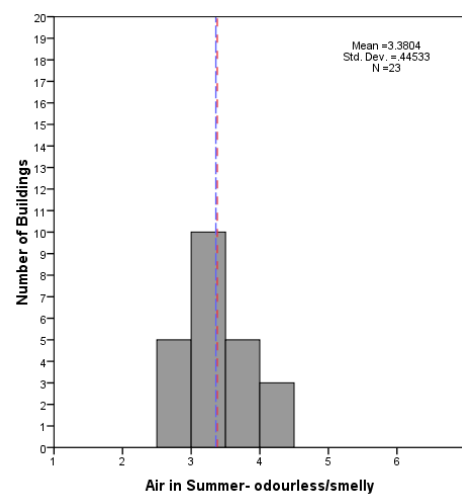


Figure 112: Air in Summer: Odourless/Smelly

Appendix H.3: Noise

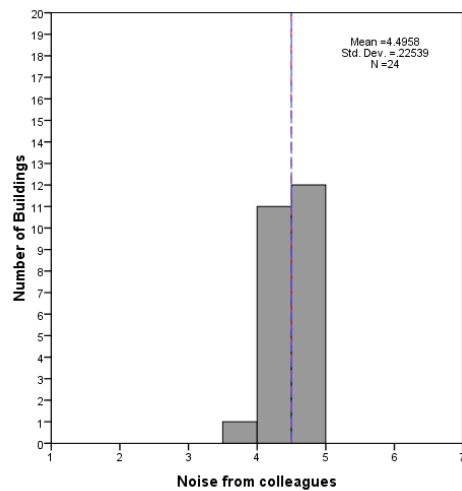


Figure 113: Noise from Colleagues

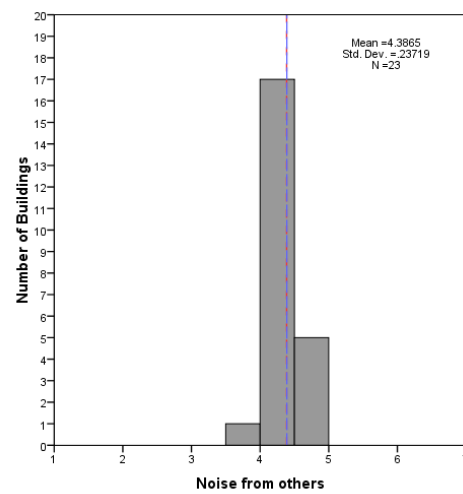


Figure 114: Noise from Others

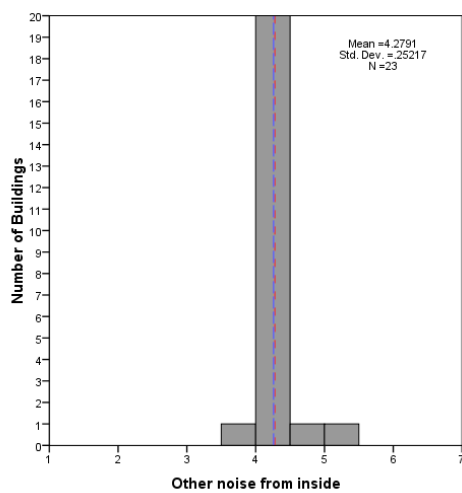


Figure 115: Other Noise from Inside

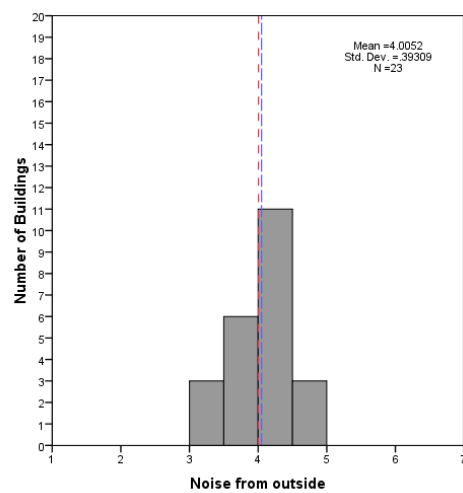


Figure 116: Noise from Outside

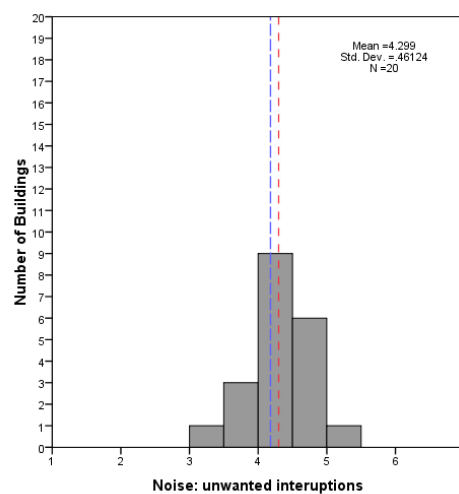


Figure 117: Unwanted Interruptions

Appendix H.4: Lighting

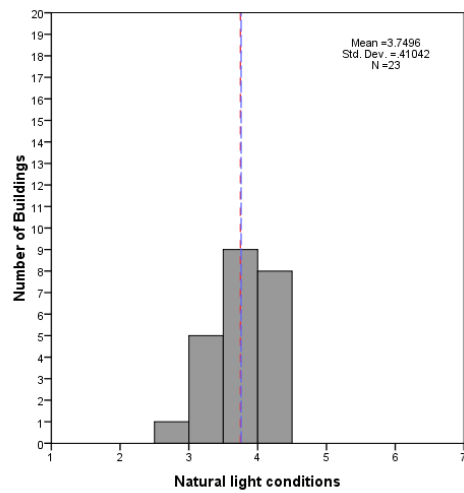


Figure 118: Natural Light Conditions

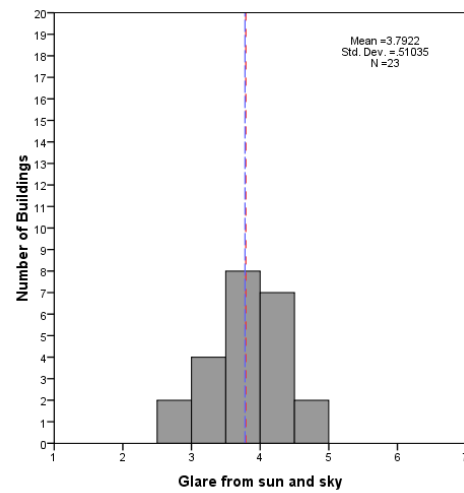


Figure 119: Glare from Sun and Sky

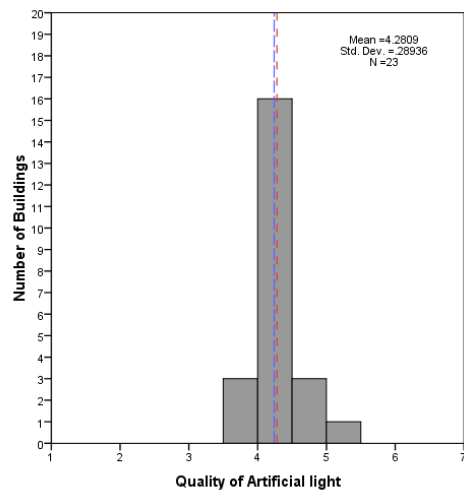


Figure 120: Quality of Artificial Light

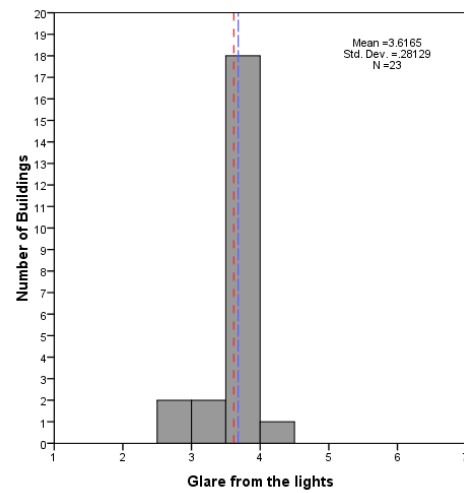


Figure 121: Glare from the Lights

Appendix H.5: Space

The only variable that was included in this section that the distribution has not already been displayed is for Space at desk. This is a 'B' scale, with an ideal score of four.

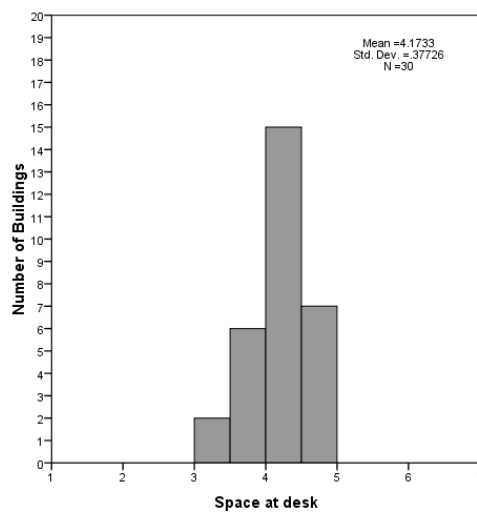


Figure 122: Space at Desk